University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations | Dissertations and Theses

March 2016

The Application of Hydrogen/Deuterium Exchange and Covalent Labeling Coupled with Mass Spectrometry to Examine Protein Structure

Nicholas B. Borotto
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

 Part of the Analytical Chemistry Commons, Biochemistry Commons, and the Structural Biology Commons

Recommended Citation

Borotto, Nicholas B., "The Application of Hydrogen/Deuterium Exchange and Covalent Labeling Coupled with Mass Spectrometry to Examine Protein Structure" (2016). *Doctoral Dissertations*. 553.
https://scholarworks.umass.edu/dissertations_2/553

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

THE APPLICATION OF HYDROGEN/DEUTERIUM EXCHANGE AND COVALENT LABELING COUPLED WITH MASS SPECTROMETRY TO EXAMINE PROTEIN STRUCTURE

A Dissertation Presented

by

NICHOLAS B. BOROTTO

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

FEBRUARY 2016

Chemistry

© Copyright by Nicholas B. Borotto 2016

All Rights Reserved

THE APPLICATION OF HYDROGEN/DEUTERIUM EXCHANGE AND COVALENT LABELING COUPLED WITH MASS SPECTROMETRY TO EXAMINE PROTEIN STRUCTURE

A Dissertation Presented

by

NICHOLAS B. BOROTTO

Approved as to style and content by:

_____
Richard W. Vachet, Chair

_____
Lila M. Gierasch, Member

_____
Igor A. Kaltashov, Member

_____
Scott C. Garman, Member

_____
Craig T. Martin, Department Head
Department of Chemistry

DEDICATION

This is for all the people whose love and support helped me get to this point.
Thank you all so much.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Richard. He has been an amazing advisor and without his guidance, patience, and insight this process would have been significantly more difficult. In addition, I want to thank Janet, John, and Lara for giving me the opportunity, the direction and the room to grow as a person and as a scientist. Working with them was instrumental in preparing me for graduate school. I would also like to thank Lynmarie, Scott, Lila, and Igor for their advice and discussions which helped both me and my work to improve and develop.

I was fortunate to have amazing labmates. They were not only always available to discuss science but they were also willing to share drawer whiskey, play pool, be targets for rubber spiders and balls, and sing songs from Moulin rouge and Disney movies. I would like to thank Adam, Yuping, Jia, Bo, Alyssa, Gokhan, Cara, Katie, Alex, Kristen, and the many others. You all made the lab an absolute pleasure to work in.

Finally I thank my family and friends for all of their love and support. I would not have been able to do this without you all.

ABSTRACT

THE APPLICATION OF HYDROGEN/DEUTERIUM EXCHANGE AND COVALENT LABELING COUPLED WITH MASS SPECTROMETRY TO EXAMINE PROTEIN STRUCTURE

FEBRUARY 2016

NICHOLAS B. BOROTTO, B.S., NORTHEASTERN UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Richard W. Vachet

Thorough insight into a protein's structure is necessary to understand how it functions and what goes wrong when it malfunctions. The structure of proteins, however, is not easily analyzed. The analysis must take place under a narrow range of conditions or risk perturbing the very structure being probed. Furthermore, the wide diversity in size and chemistry possible in proteins significantly complicates this analysis. Despite this numerous methods have been developed in order to analyze protein structure. In this work, we demonstrate that mass spectrometry (MS)-based techniques are capable of characterizing the structure of particularly challenging proteins. This is done through the study of two model systems: (1) the amyloid forming protein β2-microglobulin and (2) the protein therapeutics human growth hormone and immunoglobulin G1.

β-2-microglobulin (β2m) is an amyloidogenic protein and is the major constituent of fibrils in the disease dialysis related amyloidosis (DRA). Stoichiometric concentrations of Cu(II) have been used in vitro to induce the amyloid formation of β2m, but the structural changes caused by Cu(II) have not been fully elucidated. Other

transition metals, such as Zn(II) and Ni(II), do not cause β2m amyloid formation, yet a comparison of the structural changes caused by these metals and those caused by Cu(II) could reveal essential structural changes necessary for amyloid formation. To probe these different structural changes, we have used a combination of hydrogen-deuterium exchange (HDX) and covalent labeling together with MS. Results from these measurements reveal that Cu(II) alone is capable of inducing the *cis-trans* isomerization of the X-Pro bond of Pro32 and the other necessary conformational changes that allow β2m to form an amyloid competent state, even though Ni(II) binds the protein at the same site. We also find that Zn(II) binding leads to increased dynamics, indicating increase structural instability, which is consistent with the amorphous aggregation observed in the presence of this metal.

The second part of this dissertation investigates the use of diethylpyrocarbonate (DEPC)–based covalent labeling to detect three-dimensional structural changes in immunoglobulin G1 and human growth hormone after they have been exposed to degrading conditions. We demonstrate that DEPC labeling can identify both specific protein regions that mediate aggregation and those regions that undergo more subtle structural changes upon mishandling of these proteins. Importantly, DEPC labeling is able to provide information for up to 30% of the surface residues in a given protein, thereby providing excellent structural resolution. Given the simplicity of the DEPC labeling chemistry and the relatively straightforward mass spectral analysis of DEPC-labeled proteins, we expect this method should be amenable to a wide range of protein therapeutics and their different formulations.

In the final section of this dissertation, we demonstrate that, in certain instances, scrambling of the DEPC label from one residue to another can occur during collision-induced dissociation (CID) of labeled peptide ions, resulting in ambiguity in label site identity. From a preliminary study of over 30 labeled peptides, we find that scrambling occurs in about 25% of the peptides and most commonly occurs when histidine residues are labeled. Moreover, this scrambling appears to occur more readily under non-mobile proton conditions, meaning that low-charge state peptide ions are more prone to this reaction. For all peptides, we find that scrambling does not occur during electron transfer dissociation, which suggests that this dissociation technique is a safe alternative to CID for correct label site identification.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1
MASS SPECTROMETRY AND ITS APPLICATION TO PROTEIN STRUCTURE

1.1 Protein Structure

In order to survive and reproduce, cells must balance their osmotic pressure, maintain their structure, perform chemistry, and do an assortment of other functions. Proteins are the means by which cells perform most of these functions. Despite this wide range of functionality, proteins are simply polymers composed of 20 unique amino acids. This breadth of function is achieved by varying the identity, order, and orientation of these amino acids. Thus, in order to understand how cells function and what goes wrong when they malfunction, a thorough understanding of protein structure and function is necessary.

Amino acids are small organic molecules that contain an amine, a carboxylic acid, and often a side chain. The chemistry of these side chains varies widely, ranging from aliphatic to charged. The order and identity of these amino acids, often referred to as primary structure, is coded within the DNA of cells. For a protein to achieve its desired function, side chains that are often distal in primary structure must come closer in space. To ensure functional proteins, the folding of these polymers must occur with high fidelity.

To achieve this reproducibility, the folding of a protein is programmed into its sequence.[1] This is accomplished through the pattern in which the hydrophobic and

hydrophilic residues are intermixed in the primary sequence. The seclusion of these hydrophobic residues from water is the major driving force behind protein folding.[1,2] This process is referred to as hydrophobic collapse. Once solvent is excluded the protein must arrange in a way to satisfy the hydrogen bonding of its peptide backbone. In tightly packed regions, steric constraints limit the backbone to two main geometries capable of satisfying these bonds. These orientations result in either the formation of a helix or a strand, which are termed α-helices and β-strands. These structural elements are parts of a protein's secondary structure. Proteins are made up of multiple secondary structure components. How these components orient together to form the final three-dimensional configuration is the tertiary structure of a protein. Often understanding the primary, secondary, and tertiary structure of a protein offers insight in to its function. Sometimes, however, multiple subunits may be necessary to achieve some functions; the arrangement of these subunits is referred to as quaternary structure.

In order to fully understand how a protein achieves its function, a complete understanding of all of these aspects of structure is important. Significant research has already been dedicated to understanding the primary[3–6] and secondary[7,8] structure of proteins. While there has also been significant work dedicated to understanding the tertiary structure of proteins,[9,10] the following work will propose additional methodology to probe their tertiary and quaternary structure by examining two model systems: (1) β-2-microglobulin (β2m), and (2) a pair of therapeutic proteins (i.e. IgG1 and HGH).

1.1.1 β-2-microglobulin and Dialysis related amyloidosis

The amphiphilic character that leads to the folding of proteins can also lead to their aggregation. Aggregation is not uncommon. In fact, cells have developed a number of pathways to recover, degrade, or sequester these aggregates.[11] There are some aggregates, however, which avoid or overwhelm such rescue pathways. Amyloidosis describes a class of diseases where proteins aggregate into insoluble fibrils. These fibrous structures are resistant to these degradation pathways and their presence can be deleterious to the cell or other biological tissues. There are approximately 20 human diseases that involve this type of aggregation. The specific disease is denoted by the identity of the aggregating protein and the location of the fibril deposits.[12]

Many of these amyloidogenic proteins are natively soluble and monomeric. Despite this, the monomer is somehow stimulated into a normally unpopulated amyloidogenic conformer that can ultimately form amyloid fibrils.[13] Significant work has been dedicated to understanding and identifying the mechanisms by which these amyloidogenic conformers are populated for different proteins.[12,14–17] One event that has been shown to initiate aggregation for several of these amyloid-forming proteins is the binding of transition metals.[18–20] This interaction is thought to stabilize an amyloidogenic competent state mediating the formation of amyloids.[19–27] β-2-microglobulin (β2m) is one such amyloidogenic protein (Figure 1-1) that can form amyloid fibrils in the presence of Cu(II).

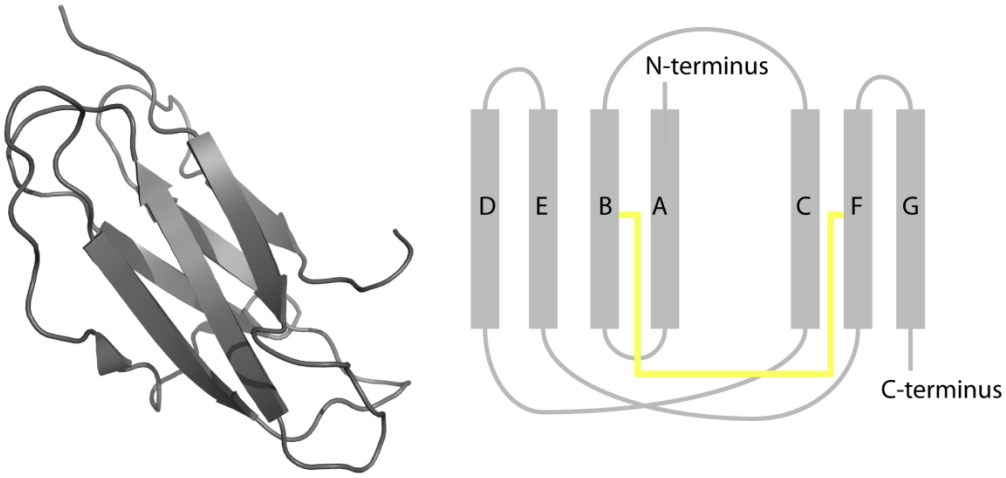

Figure 1-1 Cartoon representation of β-2-mictoglobulin and a schematic illustrating the nomenclature of its β strands (PDB: 2XKS). The yellow line represents the disulfide between Cys25 and Cys80 on the B and F strands, respectively.

β2m is normally found as a structural component of the MHC type 1 complex. Its structure has been characterized in both its native complex and its monomeric state.[28,29] β2m is a 99 residue β-sandwich protein that has an approximate molecular weight of 12,000 Daltons. It has been found to be the major constituent of fibrils in the disease dialysis related amyloidosis (DRA).[30,31] Elimination through the kidneys is the major pathway by which the body disposes of β2m. Patients in renal failure rely on dialysis to perform the function of the kidney. Dialysis, however, cannot remove β2m effectively. As a result, its concentration rises from ~0.1 μM to up to ~6 μM in the body.[32,33] In this disease the deposition of these fibrils in joints lead to acute arthropathy,[32,34] eventually necessitating joint replacement.

An abundance of research has been dedicated to understanding the process by which these fibrils are generated. Despite this, the mechanism by which they are generated *in vivo* is not known but it is known that the increased concentration of β2m

alone is not sufficient.[35,36] Prior research has also shown that acidic conditions,[37,38] certain mutations,[39,40] cleavage of the six N-terminal amino acids (ΔN6),[17,41] limited proteolysis[42], stoichiometric concentrations of Cu(II)[22,43–47], and others[48,49] have the capability to induce the amyloidosis of β2m *in vitro.*[37,50] The Cu(II) mediated pathway is particularly compelling due to the incidence of DRA decreasing significantly after the institution of Cu(II)-free dialysis membranes.[51,52]

Thus, the process by which Cu(II) binding induces fibril formation has been the subject of much study. β2m fibril formation is preceded by the formation of di-, tetra-, and hexameric species upon Cu(II) binding (Figure 1-2).[53] These oligomers, particularly the dimer[54,55] and tetramer,[55,56] were characterized and shown to maintain a native-like structure.[53,55,57,58] The effect of Cu(II) on amyloid formation was also shown to be catalytic as it is released upon a structural rearrangement of the tetramer, and the final fibrils were shown to be devoid of Cu.[57,58] The influence of Cu(II) binding on monomeric β2m has also been extensively studied, and studies indicate that Cu(II) binding induces the formation of an amyloidogenic conformer.[43,59,60]

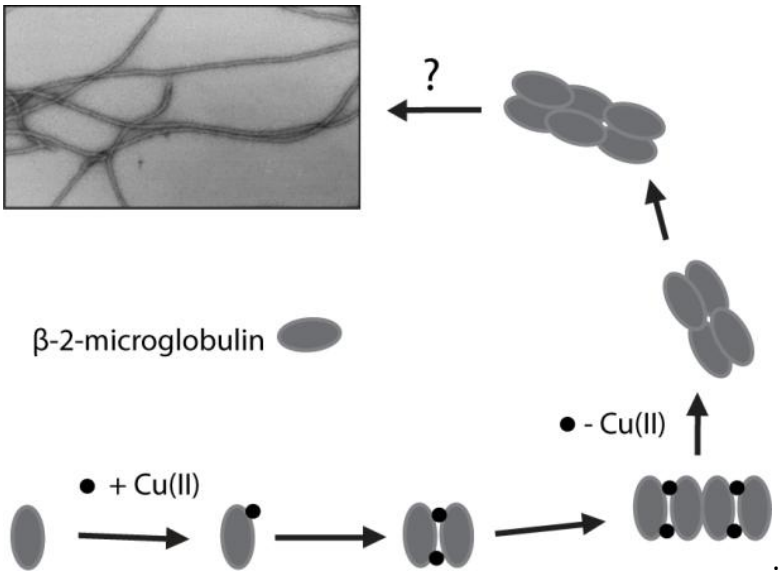

Figure 1-2 Mechanism by which β2m is converted by Cu(II) in an amyloidogenic conformer.

Suspected amyloidogenic conformers of β2m have also been studied for several mutants, including ΔN6,[61] P32G,[62] and P32A,[63] with atomic-level information being obtained by NMR or X-ray crystallography. The amyloidogenic conformer formed upon Cu(II) binding has not been completely characterized,[44] although, some of the structural changes necessary to induce the oligomerization of β2m have been revealed.[22,46,64,65] The isomerization of the His31-Pro32 amide bond, the movement of Asp59, Arg3, Phe30, and Trp60 are some of the changes that Cu(II) binding has been shown to induce.[55,64,65] A more thorough understanding of the structural changes caused by Cu that allow β2m to be transformed into amyloid aggregates may be informative for both non-Cu(II) induced transformations of β2m and the Cu(II) mediated conversions of other amyloidogenic proteins.

1.1.2 Therapeutic Proteins

Amyloid aggregation of proteins is not the only biomedically important type of protein aggregation. The unfolding and aggregation of protein therapeutics is also a growing area of interest in need of new tools for probing protein higher order structure. Protein therapeutics attempt to recover native function through the introduction of externally produced protein. Human growth hormone (HGH) and insulin were two of the earliest adopted protein therapeutics, which were used to treat hypopituitarism and diabetes mellitus type 1, respectively.[66–70] Since then, protein therapeutics have become the fastest growing segment of the pharmaceutical market with approximately 200 proteins approved for use as of 2012.[71]

Immunoglobulins, which are also known as antibodies, are a class of protein therapeutics that has been extensively utilized. These proteins are naturally found as part of the immune system where they are used to recognize and bind foreign antigens. Immunoglobulin G (IgG) is an example of one of these antibodies. It is made up of two heavy and two light chains. Each light chain is bound to a heavy by disulfide bonds and the two heavy chains are also bound together by disulfides, forming a symmetrical 150 kD protein with each half of the structure having six domains. The C_H^1, C_H^2, C_H^3, and C_L domains make up the signaling and structural part of the protein, and have primary sequences that are fairly constant across IgG molecules. Antigenic specificity is derived from the variable domains, V_H and V_L (Figure 1-3). As therapeutics, one use for IgGs is to stimulate the immune system into targeting malfunctioning cells that normally evade

recognition such as in cancer or autoimmunity.[72,73] The specificity of binding relies on the proper tertiary structure of the V_H and V_L domains.[74]

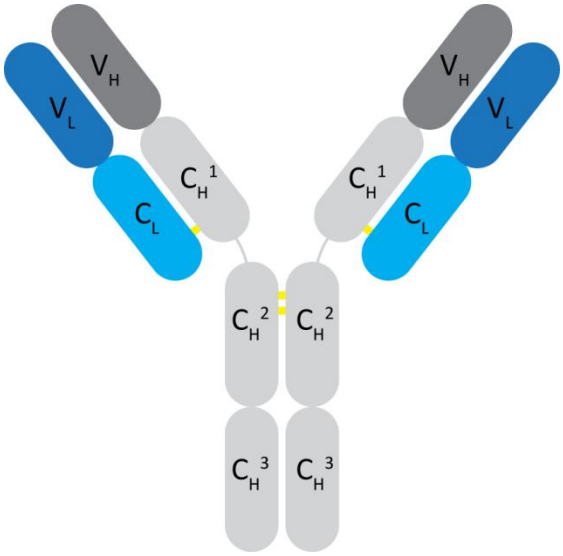

Figure 1-3 Structure of IgG1 with the domains of the heavy chain (grey) and the light chain (blue) denoted.

Perturbations in the structure of therapeutic proteins can also induce endogenously occurring antibodies to identify them as foreign. This immunogenicity can result in a reduction of the potency and efficacy of a drug and potentially lead to its complete neutralization.[75] The immunogenic response is significantly more pronounced when aggregated protein is introduced into the subject.[75] For these reasons, ensuring the proper tertiary and quaternary structure of a therapeutic protein is of the utmost importance. A thorough understanding of how protein therapeutics aggregate could allow for the targeted engineering of these molecules, leading to their improved stability.[76–78]

1.2 Tools for the Analysis of Tertiary and Quaternary Structure

The tertiary and quaternary structure of proteins is not always easily analyzed. The analysis must take place under a narrow range of conditions or risk perturbing the very structure being probed. Furthermore, the wide diversity in size and chemistry possible in proteins significantly complicates this analysis. Despite these challenges, numerous methods have been developed to analyze the structure of proteins. Therapeutic and amyloidogenic proteins specifically have been studied using Fourier transform infrared (FTIR) spectroscopy,[79–81] circular dichroism (CD) spectroscopy,[43,79,82,83] size exclusion chromatography (SEC),[22,53,57,79,84] dynamic light scattering (DLS),[57,79,85–88] fluorescence spectroscopy,[79,88–90] nuclear magnetic resonance (NMR),[87,91–96] X-ray crystallography,[97–102] and a host of other techniques. Each method, however, has its advantages and limitations.

CD and FTIR spectroscopy have been used extensively to probe the secondary structure of therapeutic and amyloidogenic proteins,[7,8,43,79–83,103–105] but they are limited in their ability to probe tertiary structure.[106,107] For example, CD is insensitive to the changes that Cu(II) binding induces in β2m.[43] Another disadvantage for these techniques is that they average all proteins and protein conformations present in solution.

SEC and DLS both measure the hydrodynamic radius of a protein, which can then be used to approximate its size. The ability to measure size is particularly useful when monitoring the soluble aggregation of proteins, making it particularly suited for the study of amyloidogenic and therapeutic proteins.[22,53,57,79,84–88] This ability to measure

size alone has limited usefulness. With β2m, for example, SEC has shown that the amyloidogenic conformer possesses a more extended structure, but more detailed structural information cannot be obtained with this method.[39]

Fluorescence spectroscopy has been used quite extensively to study protein structural changes. There are two commonly used versions of this technique that are applied to proteins: intrinsic fluorescence of tryptophan residues[108,109] and extrinsic fluorescence that relies on an added fluorophore.[110] Trp-intrinsic fluorescence has found use in the analysis of both amyloidogenic[111–113] and therapeutic[114,115] proteins. Trp fluorescence alone is unable to completely deduce the tertiary structure of proteins, but it can be used to monitor the environment of tryptophan residues allowing for some deductions about tertiary structure.[108] Extrinsic fluorescence methods can monitor a number of attributes depending on the chosen dye. Two commonly used dyes are thioflavin T (ThT) and 4,4'-bis-1-anilinonaphthalene-8-sulfonate (bis-ANS).[110] The fluorescence of ThT changes when exposed to amyloidogenic aggregates.[116,117] Consequently, the use of ThT has become one of the most common methods for identifying amyloidogenic aggregation.[14,15,38,45,88,118] The dye bis-ANS shows a strong increase in fluorescence when proteins have exposed hydrophobic surfaces[110,119] and has become widely used for measuring the structural integrity of therapeutic proteins.[79,80,114,120,121] While bis-ANS can be used to monitor the proper folding of a protein, it is unable to identify the exact location where any unfolding occurs. All of the previously mentioned techniques, while useful, have significant limitations in what can be deduced about tertiary structure.

In contrast, NMR and X-ray crystallography can be used to measure the secondary, tertiary, and quaternary structure simultaneously, thus often enabling an atomic level view of protein structure. Both techniques have been extensively used to study amyloidogenic[41,91,93,98,122,123] and therapeutic proteins,[96,100–102,124,125] including a number of IgGs.[100,101] β2m has also been analyzed extensively with these techniques, and the amyloidogenic conformers of the ΔN6 construct and the mutants P32G and P32A have been characterized using NMR[61] and crystallography,[62,63] respectively.

While these methods can provide a substantial amount of information, they also have disadvantages. Firstly, both NMR and X-ray crystallography require significant analysis time, large quantities of protein, and typically pure proteins. The analysis of large or heterogeneous proteins also remains challenging for both methods.[126–128] Thus, very few structures of intact IgG proteins have been solved.[100] The crystallographic structure of the P32A mutant displayed an increase in β-strand character along the D-strand (Figure 1-1).[63] This was initially hypothesized to be an important structural change required to initiate oligomerization. This was later discounted,[54] potentially indicating that this increase in β-strand character may merely be a crystallographic artifact. These artifacts are created due to the protein being probed in an environment that it is not found in natively. Another issue with crystallization is that the protein is crystallized in its lowest energetic state. For amyloidogenic proteins, this is its fibril form, thus limiting its applicability in amyloidogenic systems. In the case of NMR, the protein needs to be grown in isotopically-enriched media because only isotopes of nitrogen and carbon atoms are observable by NMR. NMR also has trouble with

paramagnetic metals, and so the data from Cu-binding proteins, for example, has to be carefully interpreted because of non-specific broadening of resonances. The difficulties associated with certain protein systems and limitations of both the high resolution and low resolutions methods necessitates the development of new methodologies.

1.3 Mass Spectrometric Approach to Structure

Mass spectrometry (MS)-based techniques are an alternative that could potentially fill this need. Unlike NMR and X-ray crystallography, MS is unable to probe tertiary or quaternary structure directly. Instead, MS relies on reagent molecules that can react with certain solvent-accessible sites on the protein. These modifications induce a mass change which can be measured using MS (Figure 1-4). The extent of modification can be used to measure solvent accessibility or in some cases the dynamics of the labeled site. These values and, more importantly, changes in these values give insight into changes in tertiary structure upon some perturbation. Thus, MS-based methods are best used when comparing a protein under different conditions. While MS will unlikely ever produce the same insight into structure as X-ray crystallography and NMR, it is still capable of generating complementary structural information for a wide range of protein systems.

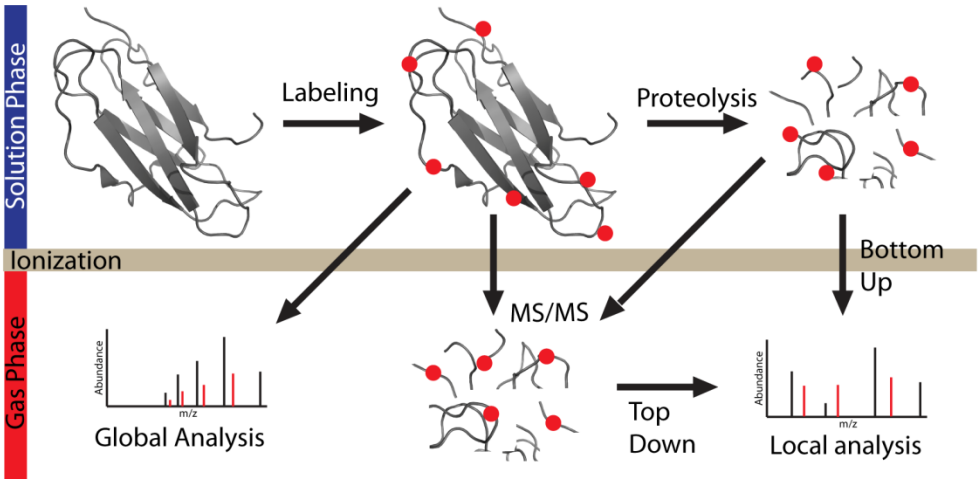

Figure 1-4 Generalized work flows for structural mass spectrometry experiments. Bottom-up (fragmentation prior to ionization) and top-down proteomics (fragmentation after ionization of intact protein) are two means of reading out the structural information that is obtained by labeling the protein in solution.

Each label site can be used as a diagnostic for changes in nearby tertiary or quaternary structure. Thus, the structural resolution of MS-based techniques is dependent on two factors: (1) the number and distribution of probe-able sites on the protein and (2) the ability to accurately assign which of those sites have been labeled. While high coverage is desirable, reagents that are capable of modifying multiple sites often require more complicated analyses. Thus, molecules with a wide range of specificities have been adopted. Identifying the location of all of these labels relies on the fragmentation of the protein into smaller pieces. The label location can be assigned through the identification of which fragments have been modified. The fragment size and the degree of overlap determine how narrowly the label location can be assigned. These fragments can either be produced prior to ionization through proteolysis in bottom-up proteomics or after using tandem mass spectrometry (MS/MS) in a top-

down proteomics approach. In many cases MS/MS can be applied to proteolysis-produced peptides to more precisely assign label locations.

1.3.1 Hydrogen/Deuterium Exchange (HDX)

In HDX/MS the amide hydrogens of the protein are exchanged with deuterium (or *vice versa*). All but one of the 20 common amino acids have an amide bond, enabling the majority of a protein's backbone to be probed. When the sample is then analyzed by a mass spectrometer as in Figure 1-4, a mass shift of one dalton will occur for every hydrogen atom that has been exchanged (Figure 1-5). The rate at which each amide hydrogen atom is exchanged is dependent on its accessibility to solvent and whether or not it is involved in hydrogen bonds. Amides involved in hydrogen bonds are protected from exchange. Hydrogen bonds, however, often form and re-form as the protein samples different conformations. Accordingly, more dynamic regions of a protein show greater amounts of deuterium exchange.[129] The ability to measure protein dynamics opposed to merely solvent accessibility differentiates this method from other MS-based approaches.

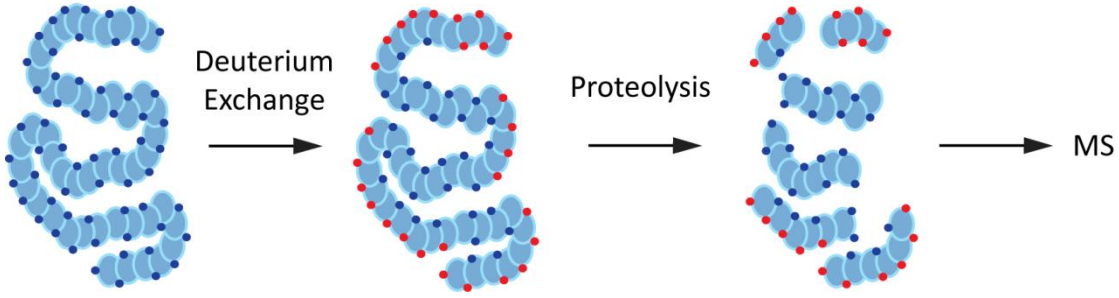

Figure 1-5 Figure illustrates amide hydrogen atoms (small blue circles) being exchanged with deuterium (red circles). Thus, allowing for regions in the protected core of the protein to be identified.

This ability to probe dynamics has led to the broad application of HDX in the analysis of protein structure by MS.[130–134] HDX/MS has also been used to probe both β2m and therapeutic proteins to a limited extent.[135–142] Global HDX studies have been used on β2m to probe the change in dynamics after its dissociation from the MHC-1 complex,[135] upon exposure to added reagents such as Cu(II),[136] SDS,[136] TFE,[136] after the removal of the six N-terminal amino acids,[136] and cleaving the protein at Lys58 via limited proteolysis.[137] These studies demonstrated that the dissociation from the MHC-1 complex, binding of Cu(II), truncation, exposure to SDS, and TFE all induce increased conformational flexibility of β2m. HDX has also found widespread use on many therapeutic proteins.[138,139] Antibodies in particular have been a focus of this research.[140] The method has been used for epitope mapping,[124,143,144] probing the effects of chemical degradation,[141] and measuring the influence of deglycosylation on the dynamics of IgG.[142]

One challenge associated with HDX/MS is the transient nature of the exchanged amide hydrogens. The lability of the exchangeable hydrogens can lead to scrambling of the added deuterium during ionization[145] or fragmentation in some tandem mass spectrometric methods.[146,147] This scrambling can lead to ambiguous data at best or completely incorrect results at worst. Another problem is back exchange. Under native conditions (e.g. pH 7 and 22°C) exchange is fast enough that any added deuterium can be lost during the proteolysis, liquid chromatography, and ionization steps of the

analysis. As a result of these issues, special care and often expensive instrumentation are required to minimize back exchange and to accurately locate deuterated sites.[132]

1.3.2 Covalent Labeling Based Methods

Methods that use covalent bond formation to characterize protein structure are not subject to back exchange. In contrast to HDX, the covalent nature of these modifications enables the use of MS/MS, often in conjunction with proteolysis, as a means to improve label assignment. Covalent labels also provide complementary information to HDX in that they report on protein side chains rather than the protein backbone.

1.3.2.1 Hydroxyl Radical Footprinting

Hydroxyl radical footprinting (HRF) is a commonly used approach for covalent labeling.[148–152] In this method, hydroxyl radicals are produced through the radiolysis or photolysis of water or hydrogen peroxide. The produced radicals then oxidize solvent accessible sites on the protein. These labels are capable of modifying numerous sites on a protein but tend to label sulfur-containing, aromatic, and aliphatic residues more rapidly than other residues.[150] Just as with HDX, the labeling reagent is roughly the size of a water molecule allowing for a good approximation of solvent accessibility.

HRF application to amyloidogenic systems is limited, and to our knowledge it has not been applied to study β2m amyloid formation. HRF has been used to probe the pre-

amyloid oligomers of Aβ40[152,153] and Aβ42.[154] The method has also recently been applied to monitor structural changes in therapeutic proteins.[83,114] The technique was shown to be quite sensitive to subtle structural changes as it was able to distinguish expired protein therapeutics from fresh ones.[83] HRF also demonstrated the ability to identify the regions of aggregation in therapeutic monoclonal antibodies (mAbs).[114]

There are some challenges associated with the implementation of HRF. Most notably, oxidation by hydroxyl radicals can produce over 50 different types of modifications that can complicate MS analysis.[150] Moreover, in its most commonly used form, a laser or synchrotron source is necessary to generate the radicals, which limits its widespread applicability.

1.3.2.2 Cross-linking

Other covalent labeling techniques use bifunctional reagents to link residues that are spatially adjacent despite being distant in primary structure. The cross-linked peptides are then sequenced and identified by MS, thereby revealing nearby residues (Figure 1-6). This method has been used to probe the structures of individual proteins[155] and protein complexes.[156–160] It has also been used to study amyloid forming and therapeutic proteins, specifically the oligomerization of Aβ[161,162] and epitope mapping of antibodies.[163] In many instances the label will only bind at a single site, thereby failing to produce a cross-link. These so called “dead ends” enable this method to probe solvent accessibility of these labeled sites, but this feature is not commonly used in protein structural analyses. One of the major drawbacks with cross-linking is that the size of

these labels may perturb the structure of the protein. Therefore, any subsequent labeling could be probing non-native structures. Another challenge associated with these bifunctional reagents is that they produce two linked peptides of various sizes each possessing an N- and C-termini. The ability of current proteomics software to identify these products is hindered by this complexity.[157]

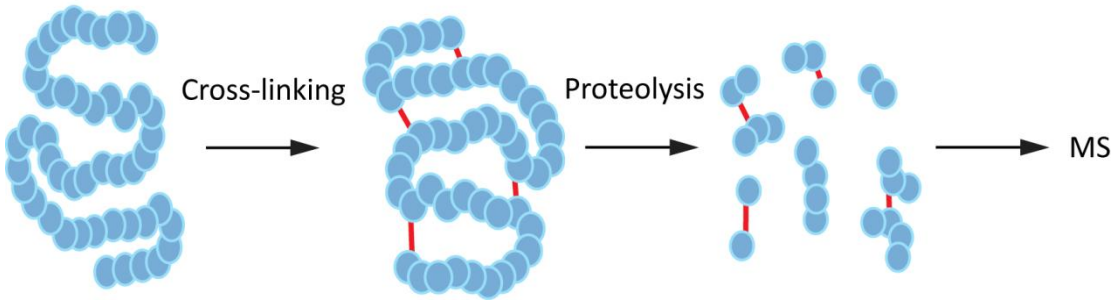

Figure 1-6 Labeling results in the linking of residues distant in primary structure but spatially adjacent due to folding. This enables some information about folding to be deduced.

1.3.2.3 Residue Specific Covalent Labeling

Another approach to covalent labeling uses amino acid-specific reagent molecules to modify solvent exposed residues.[164] The extent of modification can be used to measure solvent accessibility of each site (Figure 1-8). This approach to labeling is simple as it requires no specialized equipment and typically produces only a single type of mass addition. A wide range of reagents are available (see Figure 1-7 for examples), ranging from those that have narrow specificity (e.g. succinimides and butanedione) to those with broad specificity (e.g. diethylpyrocarbonate (DEPC)).[164] DEPC

is capable of labeling all nucleophilic residues, enabling the monitoring of approximately 30% of surface residues of the average protein.[65,165] The use of DEPC should maintain the simplicity of this type of covalent labeling while at the same time minimizing the loss of resolution compared to HDX and HRF.

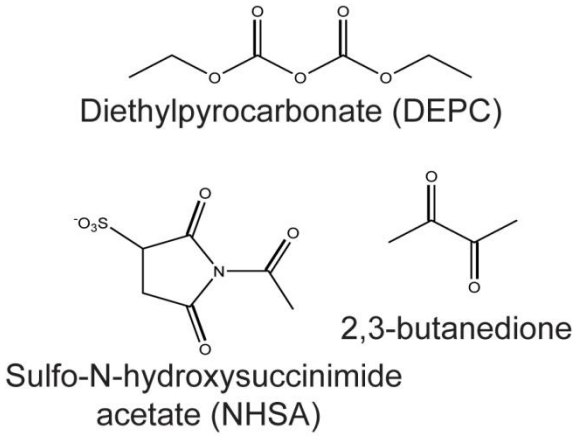

Figure 1-7 Examples of reagent molecules used for residue-specific covalent labeling.

Residue-specific covalent labeling has been used to study both amyloidogenic and therapeutic proteins, though their application to therapeutic proteins has been very limited. To our knowledge, only the carboxylate-specific reagent pair of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and glycine ethyl ester (GEE) has been used to probe the structure of a monoclonal antibody.[166] With amyloidogenic proteins, this method provides insight into protein-metal and protein-protein interactions.[65,164,165,167] For example, in β2m this method was able to both identify residues near the Cu(II) binding site and indentify residues near the interface of the preamyloid dimer.[54]

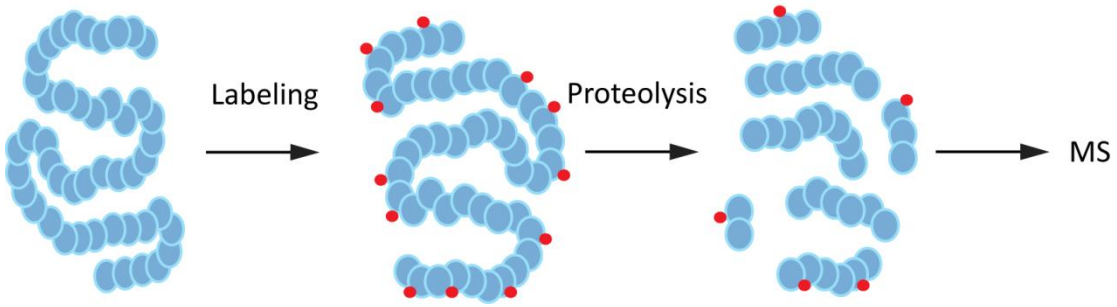

Figure 1-8 Residue specific labeling modifies solvent exposed sites enabling one to distinguish the core of the protein from its surface.

Residue-specific covalent labeling is not without drawbacks. The first is that the sizes of these reagents preclude them from sites that are normally accessible to water. Consequently, this method does not strictly probe solvent accessibility. Moreover, labeling with such large labels may perturb the structure of the protein. Thus, any subsequent labeling could potentially be probing non-native structures. To ensure this is not happening a number of methods have been developed: (1) limiting the extent of labeling to approximately one label per protein, (2) monitoring the structure of the protein through complementary biophysical techniques, and (3) ensuring the second-order character of the labeling reaction.[65,164]

All current methodologies that probe the three-dimensional structure of proteins have limitations. While MS-based approaches are no exception, they provide complementary structural information, require little sample, and are capable of probing challenging proteins and protein systems. Despite these potential advantages the application of these MS-based methods has not been fully explored.

The following sections of this dissertation will investigate and discuss the application of HDX and residue specific covalent labeling to two model systems. The first system to be discussed in chapter 2 is the amyloid forming protein β-2-microglobulin. Here we investigate the structural changes induced by copper that make it unique among metals for inducing amyloid formation. In chapter 3 we will investigate the ability of residue specific covalent labeling to distinguish changes in therapeutic proteins after exposure to denaturing conditions. Chapter 4 describes a gas phase phenomenon where, under particular conditions, certain residue-specific labels can migrate to different residues within a peptide. The final chapter will discuss the impact of this work and its future directions.

CHAPTER 2
CU(II)'S UNIQUE ABILITY TO INDUCE B-2-MICROGLOBULIN'S AMYLOIDOGENIC AGGREGATION

2.1 Introduction

Protein amyloids are a special class of protein aggregates that are associated with several human diseases in which proteins aggregate into insoluble fibrous structures that are resistant to degradation. There are about 20 human diseases that involve this type of aggregation and in each case the identity of the protein and the location of fibril deposition are different. β-2-microglobulin (β2m) is a one such amyloidogenic protein. It is naturally found as a structural component of the MHC type 1 complex, but in people undergoing dialysis treatment as a result of kidney disease, β2m forms amyloid fibrils, which are the main pathology of dialysis related amyloidosis (DRA).[30,31] These fibrils congregate in the joints and induce acute arthropathy.[32,34] Amyloid formation by β2m is initiated by the inability to eliminate β2m effectively, resulting in an increase in serum concentration from ~0.1 μM to up to ~6 μM.[32,33] An abundance of research has been dedicated to understanding the process by which these fibrils are generated. Despite this, the mechanism by which they are generated *in vivo* is not known, but it is known that increased concentrations of β2m alone are not sufficient.[35,36] Research has also shown that acidic conditions,[37,38] certain mutations,[39,40] cleavage of the six N-terminal amino acids,[17,41] limited proteolysis,[42] stoichiometric concentrations of Cu(II),[22,43–47] and others[48,49] can induce the amyloidosis of β2m *in*

vitro.[37,50] Our group has become interested in the Cu(II) mediated pathway due to the decreased incidence of DRA upon the use of Cu(II)-free membranes,[51,52] and because Cu(II) is a convenient means of triggering β2m amyloid formation in vitro.

In previous work by our group and others, it has been found that β2m fibrils are preceded by the formation of soluble di-, tetra-, and hexameric species upon Cu(II) binding (Figure 1-2).[47,53] The oligomers, particularly the dimer[54,55] and tetramer[55,56] maintain a native-like structure.[53,55,57,58] The binding of Cu(II) has been shown to play a catalytic role in the formation of β2m fibrils as it is necessary for oligomer formation but is released before the final fibrils are formed.[57,58] The influence of Cu(II) binding on monomeric β2m has also been extensively studied and it has been demonstrated that Upon binding to the monomeric β2m, Cu(II) induces several structural changes that are necessary for the formation of the dimer. These structural changes include the *cis-trans* isomerization of the His31-Pro32 amide bond that causes a repacking of the hydrophobic core and the repositioning of Arg3 and Asp59 to enable the formation of dimer-stabilizing salt bridges.[22,55,63]

Recent work has also explored how different metals influence β2m aggregation, including the similar-sized metals Zn(II) and Ni(II).[22] Interestingly, Zn(II) binding causes amorphous aggregation, while Ni(II) binding failed to initiate any aggregation. The latter observation is particularly intriguing as Ni(II) and Cu(II) both bind to the N-terminus and His31.[22] It has been suggested that metal binding to His31 lowers the barrier for the *cis-trans* isomerization of the His-Pro32,[63] so the failure of Ni(II) to cause amyloid formation is a curious observation. In this work we set out to elucidate the amyloid-causing

structural changes initiated by Cu(II) binding that evidently do not occur upon Ni(II) and Zn(II) binding. Our results show that only Cu(II) is able to cause the necessary Pro32 isomerization, but we also find other subtle structural changes that are necessary for the formation of β2m amyloids. Overall, our results highlight the fact that very specific conformational changes are necessary to achieve the amyloid-competent state of β2m.

2.2 Experimental Procedures

2.2.1 Materials

Diethylpyrocarbonate (DEPC), dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB), deuterium oxide, pepsin, imidazole, 3-morpholinopropanesulfonic acid (MOPS), potassium acetate, potassium bromide, urea, zinc sulfate, deuterium oxide, and dithiothreitol (DTT) were obtained from Sigma-Aldrich (St. Louis, MO). Urea was purchased from Mallinckrodt Chemicals (Phillipsburg, NJ). Trypsin and chymotrypsin were purchased from Promega (Madison, WI) and Roche Diagnostics (Indianapolis, IN) respectively. Tris(hydroxymethyl)-aminomethane (Tris) and tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) were purchased from EM Science (Gladstone, NJ). Human β2m purified from human urine was purchased from Lee Biosolutions (St. Louis, MO). Ammonium acetate, methanol, acetonitrile, glacial acetic acid, copper sulfate, and nickel sulfate were obtained from Fisher Scientific (Fair Lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from

Millipore (Burlington, MA). Deionized water was prepared from a Millipore (Burlington, MA) Simplicity 185 water purification system.

2.2.2 Sample Preparation

For HDX experiments a stock of 4.1 mM β2m was made in 25 mM MOPS and 150 mM potassium acetate at pH 7.4. All stocks were made fresh daily. For HNSB labeling experiments a 75 μM solution of β2m was prepared in 150 mM potassium acetate, and 25 mM MOPS (pH 7.4). For DEPC reactions 100 μM β2m in 200 mM potassium acetate, 500 mM urea, and 25 mM MOPS (pH 7.4) at 37 °C. All components were equilibrated at 22 °C prior to metal addition, and immediately returned to 37 °C after mixing. In all experiments the following metal to β2m ratios were used: Cu 2:1, Ni 16:1, and Zn 4:1. These ratios were chosen to ensure that the metal was 95% bound based on previous K_d measurements.[44]

2.2.3 Hydrogen/Deuterium exchange (HDX)

Concentrated stocks of β2m, potassium acetate and MOPS buffer, and the desired metal salt were all made and diluted into D_2O simultaneously. For all HDX experiments, the resulting concentrations were 75 μM β2m, 25 mM MOPS 150 mM potassium acetate and either 300 μM Zn, 1200 μM Ni, or 150 μM Cu. The total volume of the reaction mixture was 55 μL. The samples were then allowed to incubate in deuterium for 60, 80, 120, 240, 420, 660, 1200, 2400, 5400, or 10800 sec after which the

sample was placed on ice for an additional 30 sec. The exchange reaction was then stopped by lowering the solution pH to 2.5 using a solution of formic acid that also contained 100 mM TCEP. The total volume after quench was 110 µL. The sample was allowed to sit on ice for 1.5 min prior to proteolysis with pepsin.

2.2.4 Covalent Labeling

Two types of covalent labeling reactions were conducted. The first used HNSB to modify solvent exposed Trp residues.[164] Stock solutions of HNSB were prepared in water. Labeling of β2m by HNSB was performed for 45 sec at 22 °C and was initiated through the addition of 68.5 µM of HNSB. The total reaction volume was typically 27 µL. The HNSB labeling reaction was quenched through the addition of 10 mM tryptophan.

The second type of covalent labeling reaction involved the reagent DEPC, which can label solvent exposed His, Tyr, Thr, Ser, and Lys residues.[54,55,65,164] Stock solutions of DEPC were prepared in acetonitrile. DEPC labeling of β2m was performed for 1 min at 37 °C and was initiated by adding 0.25 mM DEPC. The total reaction volume for the experiments was 30 µL, and the total amount of acetonitrile added was < 1.5 %. The reactions were quenched after 1 min by adding 5 mM imidazole.

2.2.5 Proteolytic Digestion

β2m samples that underwent HDX were digested using pepsin. The digestion was initiated through the addition of ~1.9 µM pepsin to the already quenched samples,

resulting in a 1:20 pepsin/β2m ratio. The digestion was allowed to proceed on ice for 6 min. The digested samples were then immediately analyzed by LC/MS.

DEPC labeled β2m samples were purified using a 10,000 MWCO filter and reconstituted with 25 mM Tris-HCl (pH 7) and 1 mM $CaCl_2$ to a final concentration of 300 µM. Purified β2m samples were first reacted with 10 mM DTT at 37 °C for 45 min to reduce the disulfide bonds. The reduced protein samples were then unfolded in 12% acetonitrile at 37 °C for 45 min. Trypsin or chymotrypsin (1 µg/µL) was then added to the modified samples to yield a final enzyme/substrate ratio of 1:20. All samples were digested at 37 °C for 16 h before inactivating the enzymes by the addition of 2 µL of acetic acid. The samples were then immediately analyzed by LC/MS.

2.2.7 HPLC Separation

To analyze the digests from both the HDX and covalent labeling experiments, an HP1100 (Agilent, Wilmington, DE) HPLC system with a C18 column (15 cm x 2.1 mm, 5 µm particle size) from Supelco, (St. Louis, MO) was used. A 5 µL injection loop and an injection volume of 7 µL were used for all replicates. Peptic fragments of the H/D exchanged samples were eluted at a flow rate of 0.25 mL/min using a gradient of acetonitrile containing 0.1% formic acid that increased 1 to 23% during the first minute, 23 to 24% from 2 to 8 minutes, and then from 24 to 99% from 8 to 11 minutes. The remaining percentage of the mobile phase was water with 0.1% formic acid. The LC separation for HDX was performed on ice to minimize back exchange. The DEPC-modified proteolytic fragments were separated using a linear gradient of methanol with

0.1% acetic acid that increased from 5 to 70% over 30 min and 70 to 100% over the final 3 min. The remaining percentage of the mobile phase was water with 0.1% acetic acid.

Some of the HNSB-labeled β2m samples were analyzed as intact proteins. To do this, the samples were first desalted using a Thermo Scientific Ultimate 3000 HPLC system (Thermo Scientific, Tewksbury, MA) fitted with a Protein MicroTrap (Michrom, Auburn, CA). A 5 µL injection loop and an injection volume of 7 µL were used for all replicates. The protein was eluted using an acetonitrile gradient that increased from 1 to 99% over 5 min at a flow rate of 4 µL/min.

2.2.9 Mass Spectrometry

Mass spectral analyses of the HPLC separated samples from the HNSB labeling experiments and the HDX reactions were acquired on a Bruker Amazon ETD (Billerica, MA) quadrupole ion trap mass spectrometer equipped with an electrospray ionization source. The electrospray source conditions, including the voltage and temperature, were chosen to optimize the peptide signal. Tandem mass spectra were acquired using collision-induced dissociation (CID) with isolation widths of 1.0 Da and excitation voltages between 0.6 and 1.0 V.

The DEPC-labeled β2m samples were analyzed using a Bruker Esquire-LC quadrupole ion trap mass spectrometer (Billerica, MA) equipped with an ESI source. The ESI source was operated at a spray voltage of 3.5 kV, and the capillary temperature was set at 300 °C. The voltages for the transfer optics between the ESI source and the ion trap were optimized for maximum signal, with typical skimmer 1 and capillary offset

values of 30-35 V and 50-60 V, respectively. Tandem mass spectra were acquired using CID with isolation widths of 1.0 Da and excitation voltages between 0.6 and 1.0 V.

2.2.10 Solvent Accessibility Calculations

To calculate the solvent accessibility of surface residues, the online tool GetArea[168] was used on the intact (pdb: 2xks) and truncated (pdb:2xku) structures of β2m solved that were described by Eichner *et al.*[61]

2.3 Results

Amyloid formation by β2m is thought to be initiated by the *cis-trans* isomerization of the His31-Pro32 amide bond. Evidence for this conformational switch has come from studies of β2m in the presence of acid and Cu(II) as well as from results from the ΔN6 mutant.[61,63,64] In each instance, the isomerization of the His31-Pro32 amide bond induces a repacking of the hydrophobic core[50,53,60,62,63,112] and changes in the hydrogen bonding between the B, E, and D β strands. A comparison of the NMR structures of wild-type protein and the ΔN6 mutant reveal these structural changes and provide a model for the amyloidogenic conformer (Figure 2-1 and Table 2-1). Because Ni(II) and Zn(II) are unable to stimulate β2m amyloid formation, whereas Cu(II) can, we hypothesize that Ni(II) and Zn(II) do not enable β2m populate the amyloidogenic state. To test the hypothesis, we used HDX/MS and covalent labeling/MS to probe the structural changes caused by metal binding and compared them to changes that are expected for the

amyloidogenic conformer. Changes in hydrogen bonding (e.g. fraying between the B, E, and D β strands) can be probed using HDX and repacking of the hydrophobic core, most notably the burial of Trp60, can be probed using the tryptophan specific label dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromine (HNSB) (Figure 2-2).[164,169,170]

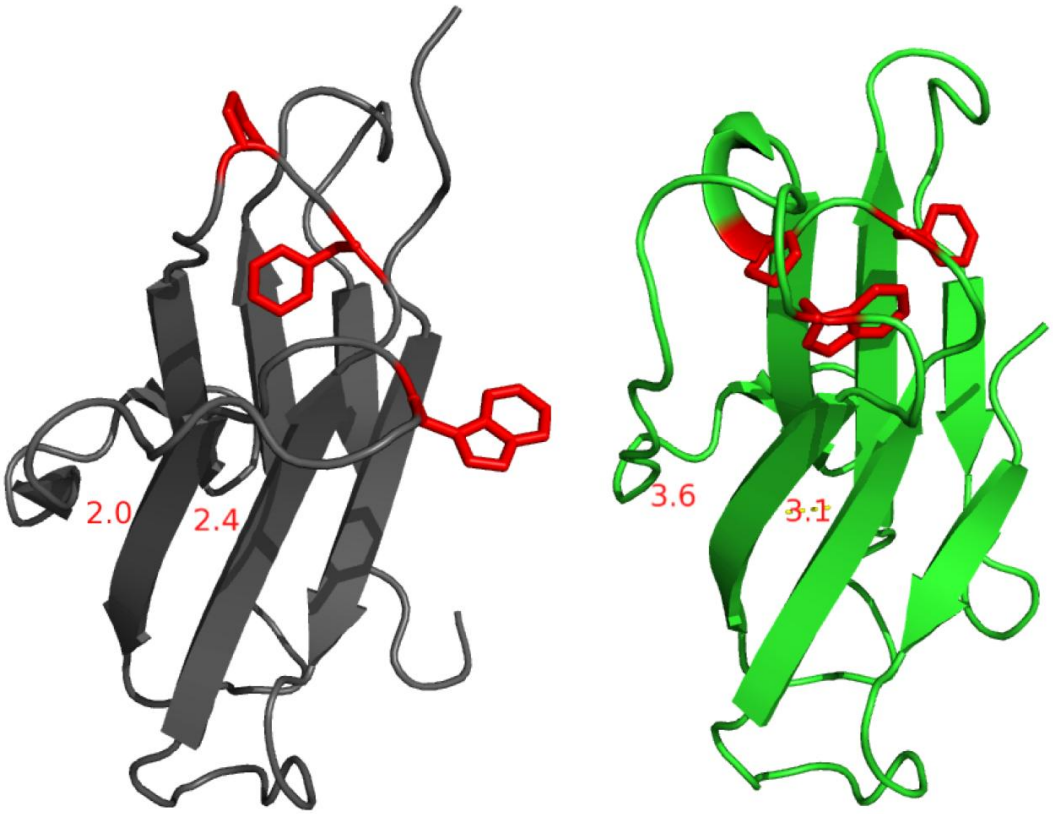

Figure 2-1 NMR structures of β2m (grey, PDB:2XKS) and ΔN6 (green, PDB:2XKU) The burial Trp60, the exposure of Phe30, and isomerization of the x-pro bond of Pro32 are illustrated as sticks in both structures. The numbers on the figure illustrate the distance between the D- and E-(left) and the E- and B- (right) strands in angstroms.

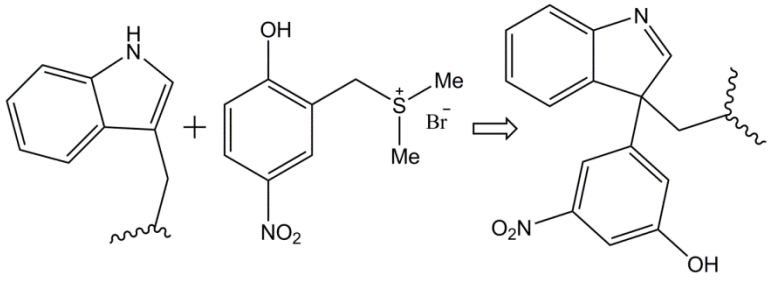

Figure 2-2 Scheme of NHSB labeling a tryptophan. Modification leads to a mass change of 151 Daltons.

Table 2-1 Observable structural changes that occur during the conversion of β2m from its native state to its amyloidogenic conformer

N-terminus	Displaced
Pro32	Isomerization
Phe30	Exposure
Phe62	core repacking
Phe56	core repacking
Tyr63	core repacking
Trp60	Burial
B-E strand	Fraying
D-E Strand	Fraying
Surface residues	Changes in solvent accessibility
Secondary Structure	Mostly intact

2.3.1 Cu-β2m vs. β2m

Cu has been shown to initiate the amyloid formation of β2m by catalyzing the *cis-trans* isomerization of the His31-Pro32 amide bond.[63] Evidence for the consequent structural changes can be found in both HDX and covalent labeling studies. A comparison of the HDX/MS results (Figure 2-3) from β2m with and without Cu(II) reveal increased amide exchange around Tyr66 and Tyr67, just as would be expected from the fraying of the B, E, and D β strands (Figure 2-3 Table 2-1). Increased amide exchange would also be expected around Cys25 and His51. While this increased exchange is evident at longer exchange times, the large proteolytic fragments (10-25 and 40-55) that are obtained also include the A-B loop and C-D loop. Thus, these fragments also have 9 and 11 unprotected amide hydrogens, respectively, causing the change to be less clear on these rapidly exchanging fragments (Figure 2-4). HDX in the remainder of the protein's backbone remains essentially the same with and without Cu(II), which is

consistent with most of the protein's backbone retaining its native-like structure upon converting to the amyloidogenic state (Table 2-1).

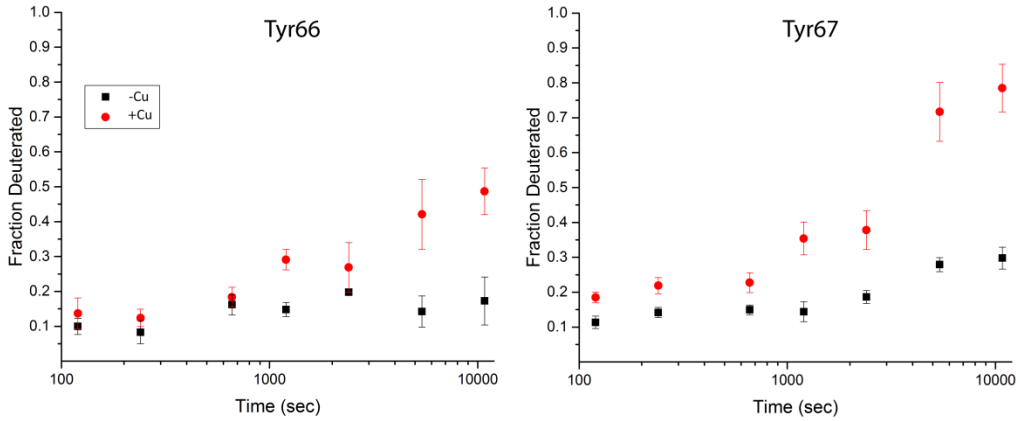

Figure 2-3 Extent of HDX at Tyr66 and Try67 without and with Cu(II). These graphs are obtained through the subtraction of three overlapping fragments 63-65, 63-66, and 63-67. These are indicative of increased dynamics between both the B- and E- β strands and the D- and E- β strands.

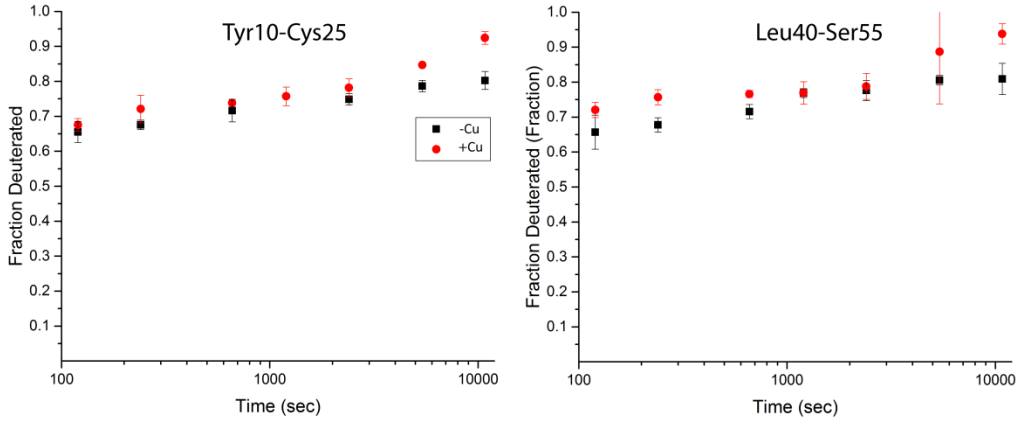

Figure 2-4 Extent of HDX for the peptic fragments 10-25 and 40-55 without and with Cu(II). Because these fragments contain unstructured regions along with β strands, changes in the extent of exchange are somewhat obscured by the fast exchanging unstructured regions.

Another key difference between the native protein and its amyloidogenic state is the burial of Trp60 upon *cis-trans* isomerization of His31-Pro32 amide bond. To probe the extent of Trp60 burial upon Cu(II) binding, we used the Trp-specific reagent HNSB,

which reacts with solvent exposed Trp residues. β2m has two Trp residues (e.g. Trp60 and Trp95), but only Trp60 is solvent exposed. Upon using a 45 sec pulse of HNSB, we find that labeling at Trp60 remains essentially unchanged for up to 60 min in the absence of Cu(II) (Figure 2-5). When Cu(II) is present, however, Trp60 undergoes a time-dependent decrease in labeling that is consistent with burial of Trp60 over time as the amyloidogenic state is populated (Figure 2-5). Trp95 also undergoes a slight increase in labeling efficiency when Cu(II) is present, which is consistent with it becoming more exposed in the amyloidogenic state. An interesting aspect of the decrease in Trp60 labeling over time is that it seems to indicate the rate at which the amyloidogenic state is formed after Cu(II) is added. Future work will explore this further.

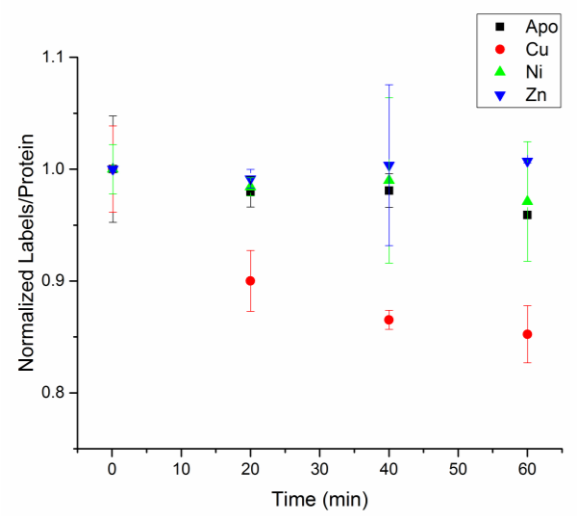

Figure 2-5 HNSB labeling of intact β2m in the presence and absence of metals. The absence of metal and the presence of Zn(II) or Ni(II) induce no change in labeling with 60 minutes. The presence of Cu(II), however, causes a decrease in labeling over time, which is consistent with the burial of Trp60 upon formation of the amyloidogenic state.

Another characteristic of the amyloidogenic state is the decreased solvent accessibility of several residues in addition to Trp60 (Table 2.1). A fraction of these

residues, which are highlighted in Table 2-2 can be probed by covalent labeling with DEPC. A comparison of the reactivity of these residues in the presence and absence of Cu(II) reveals that most of these sites follow the expected trend upon Cu(II) binding (Table 2-3). For example, Lys19, and Lys75, which are expected to undergo a 25%, and 15% decrease in solvent accessibility, respectively, each show a drop in labeling efficiency. It should be noted, though, that not all residues undergo the expected change in reactivity (e.g. Tyr67). This fact might be due to subtle, but important, differences between the Cu-induced amyloidogenic state and the amyloidogenic structure adopted by the ΔN6 mutant used as the model system.

Table 2-2 Atom specific solvent accessibility for intact β2m (2XKS), ΔN6 (2XKU), and the change between the structures for DEPC probable atoms

#	Atom	Residue	Solvent accessibility (Å^2)			#	Atom	Residue	Solvent accessibility (Å^2)		
			2XKS	2XKU	Change				2XKS	2XKU	Change
1	N	ILE 1	32.1			867	OG	SER 55	6.1	20.7	14.6
65	OG1	THR 4	23.7			898	OG	SER 57	17.8	13.9	-3.9
96	NZ	LYS 6	37.8			912	NZ	LYS 58	30.8	37.4	6.7
173	OH	TYR 10	21.0	32.2	11.2	967	OG	SER 61	9.7	10.2	0.6
188	OG	SER 11	4.6	1.5	-3.1	1004	OH	TYR 63	17.3	27.0	9.7
224	ND1	HIS 13	10.9	2.0	-9.0	1063	OH	TYR 66	12.4	13.1	0.7
227	NE2	HIS 13	15.3	13.5	-1.7	1084	OH	TYR 67	26.4	34.3	8.0
303	NZ	LYS 19	34.7	26.2	-8.5	1099	OG1	THR 68	0.9	0.0	-0.9
322	OG	SER 20	21.5	27.1	5.7	1148	OG1	THR 71	17.1	12.4	-4.7
416	OH	TYR 26	29.7	31.6	1.9	1176	OG1	THR 73	1.1	0.8	-0.4
447	OG	SER 28	2.7	21.7	18.9	1208	NZ	LYS 75	40.3	34.5	-5.7
486	ND1	HIS 31	6.2	6.7	0.6	1260	OH	TYR 78	5.6	5.2	-0.3
489	NE2	HIS 31	11.4	10.1	-1.3	1350	ND1	HIS 84	0.0	0.0	0.0
516	OG	SER 33	15.9	17.4	1.5	1353	NE2	HIS 84	2.1	2.4	0.4
642	NZ	LYS 41	29.0	14.2	-14.8	1382	OG1	THR 86	0.0	28.1	28.1
758	NZ	LYS 48	35.7	39.6	3.8	1415	OG	SER 88	20.0	26.9	7.0
809	ND1	HIS 51	3.3	0.6	-2.7	1460	NZ	LYS 91	36.6	29.9	-6.7
812	NE2	HIS 51	17.3	11.8	-5.4	1517	NZ	LYS 94	16.2	35.9	19.7
825	OG	SER 52	19.1	27.6	8.5						

Table 2-3 Modification percentages for each modified residue in the absence and presence of Cu(II), Ni(II), or Zn(II)

Residue	Change in SASA (Å^2)	apo-β2m	Cu(II)	Ni(II)	Zn(II)
His13	-1.7	47 ± 2	46 ± 3	46 ± 2	46 ± 3
Lys19	-8.5	22 ± 1	19 ± 2	18.0 ± 0.7	24 ± 2
Ty26	1.9	0.75 ± 0.07	0.79 ± 0.05	1.10 ± 0.05	0.75 ± 0.03
Ser28$^{\alpha}$	18.9	0.24 ± 0.01	0.17 ± 0.02	0.35 ± 0.02	0.26 ± 0.01
His31$^{\alpha}$	-1.3	1.8 ± 0.1	0.7 ± 0.1	1.6 ± 0.1	1.40 ± 0.05
Ser33$^{\alpha}$	1.5	1.7 ± 0.1	1.1 ± 0.1	1.11 ± 0.06	1.6 ± 0.1
His51	-5.4	61 ± 3	61 ± 4	72 ± 2	81 ± 4
Ser57/Lys58	6.7	43 ± 2	42 ± 3	19.9 ± 0.8	46 ± 2
Tyr63$^{\beta}$	9.7	7.1 ± 0.5	4.3 ± 0.3	3.6 ± 0.3	5.4 ± 0.3
Tyr67	8.0	2.1 ± 0.1	1.75 ± 0.06	2.4 ± 0.1	2.01 ± 0.08
Lys75	-5.7	0.36 ± 0.04	0.33 ± 0.02	0.35 ± 0.04	0.52 ± 0.04
Ser88	7.0	63 ± 3	65 ± 3	60 ± 1	74 ± 1
Lys94	19.7	30 ± 2	29 ± 2	39 ± 3	39.7 ± 0.5

$^{\alpha}$ Labeling may be influenced due to proximity to Cu(II) binding site.
$^{\beta}$ Labeling may be influenced by the binding of Cu by Asp59.

Changes in the HDX and covalent labeling behavior of β2m with and without Cu(II) at the expected locations provide a benchmark against which the Zn(II) and Ni(II) bound proteins can be compared. This comparison enables us to determine whether Zn(II) and Ni(II) induce the amyloidogenic conformer.

2.3.2 Ni-β2m vs. β2m

The influence of Ni(II) binding on β2m is particularly interesting as it binds in a very similar location to Cu(II), yet it induces no oligomerization or aggregation.[22] When the Ni(II)-bound form of β2m is investigated and compared to the metal free protein using HDX/MS, we find no increase in exchange at Tyr66, suggesting that the hydrogen bonding between the B and E strands is unchanged (Figure 2-6). This lack of change is in stark contrast to the Cu(II) bound form (Figure 2-3). A slight increase in dynamics between the D and E strands (Figure 2-6, Tyr67 data) is observed, but this increase is extensive than when Cu(II) is bound to the protein. The complementary proteolytic fragments (10-25 and 40-55) do not show any significant increase in exchange upon Ni(II) binding (Figure 2-7). Overall, the HDX results with Ni(II)-bound form of the protein suggest that Ni(II) binding does not induce the same structural change as Cu(II) binding.

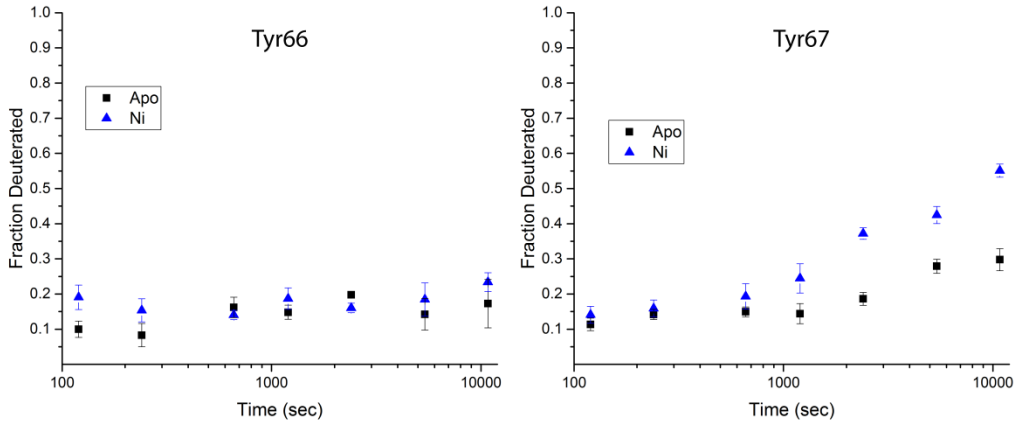

Figure 2-6 Extent of HDX at Tyr66 and Try67 without and with Ni(II). These graphs are obtained through the subtraction of three overlapping fragments 63-65, 63-66, and 63-67. These are indicative of increased dynamics between both the B- and E- β strands and the D- and E- β strands.

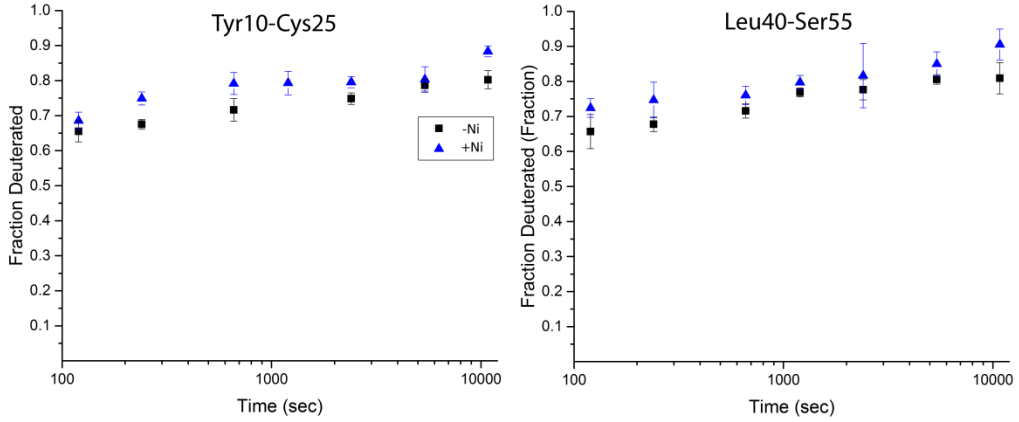

Figure 2-7 Extent of HDX for the peptic fragments 10-25 and 40-55 without and with Ni(II). Because these fragments contain unstructured regions along with β strands, changes in the extent of exchange are somewhat obscured by the fast exchanging unstructured regions.

Changes in the solvent accessibility of Trp60 upon binding Ni(II) were also assessed using HNSB labeling. As indicated in Figure 2-5, the metal-free protein and the Ni(II)-bound protein undergo no change in HNSB labeling over a period of 60 min. Covalent labeling with DEPC also suggests different structural changes upon Ni(II) binding (Table Y). DEPC labeling of the Ni(II)-bound protein shows significant deviations from the Cu(II)-bound form at Tyr26, Ser28, His51, Lys58, Tyr67, and Lys94. The full

consequences of these changes are not clear, but these data re-emphasize the intriguing observation that Cu(II) and Ni(II) induce different structural changes despite both having the N-terminus and His31 as common binding sites. The fact that Trp60 is not buried upon Ni(II) binding, together with minimal changes in HDX behavior, strongly suggests that NI(II) binding does not induce the *cis-trans* isomerization of Pro32. Evidently Ni(II) binding to His31 is not sufficient to induce the isomerization that is observed with Cu(II). Furthermore, the fact that Ni(II) binding does induce a slight increase in amide exchange between the D and E strands, yet does not induce oligomerization or aggregation, indicates that the destabilization of the D strand alone is not enough to induce aggregation.

2.3.3 Zn-β2m vs. β2m

The HDX behavior of the Zn(II)-bound form of β2m is notably different than the exchange behavior of the Cu(II)-bound form. Changes in amide exchange induced by Zn(II) binding are evident even at the first measured time point at many different site in the protein (Figures 2-8, 2-9 and 2-10), possibly signifying a decrease in protein stability. Increased exchange in the presence of Zn(II) is consistent with the fact that Zn(II) binding destabilizes β2m by as much as 14 kJ/mol.[44] Like the Ni(II)-bound form of β2m, Zn(II) fails to induce an increase in exchange between the B and E strands, despite the overall increase in protein dynamics. Unlike Ni(II), however, Zn(II) does increase exchange at the D and E strands (Figure 2-8) and near the A-B loop (Figure 2-9). Even

though Zn(II) does cause an increase in the dynamics at the D and E, the increase in exchange is less than that observed upon Cu(II) binding.

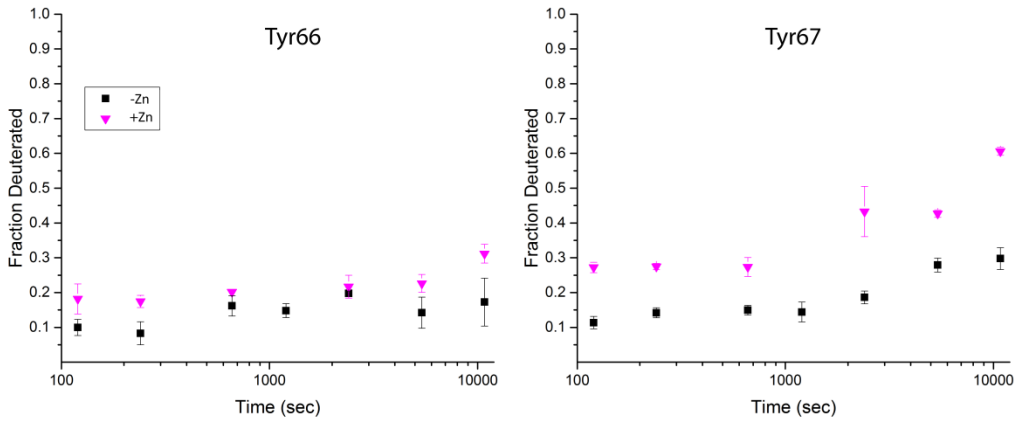

Figure 2-8 Extent of HDX at Tyr66 and Try67 without and with Zn(II). These are indicative of increased dynamics between both the B- and E- β strands and the D- and E- β strands. These graphs are obtained through the subtraction of three overlapping fragments 63-65, 63-66, and 63-67.

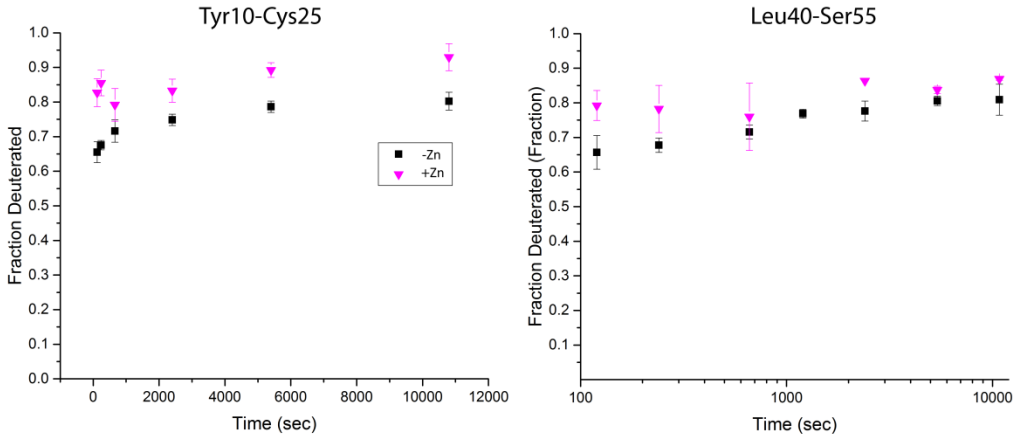

Figure 2-9 Extent of HDX for the peptic fragments 10-25 and 40-55 without and with Zn(II). Because these fragments contain unstructured regions along with β strands, changes in the extent of exchange are somewhat obscured by the fast exchanging unstructured regions.

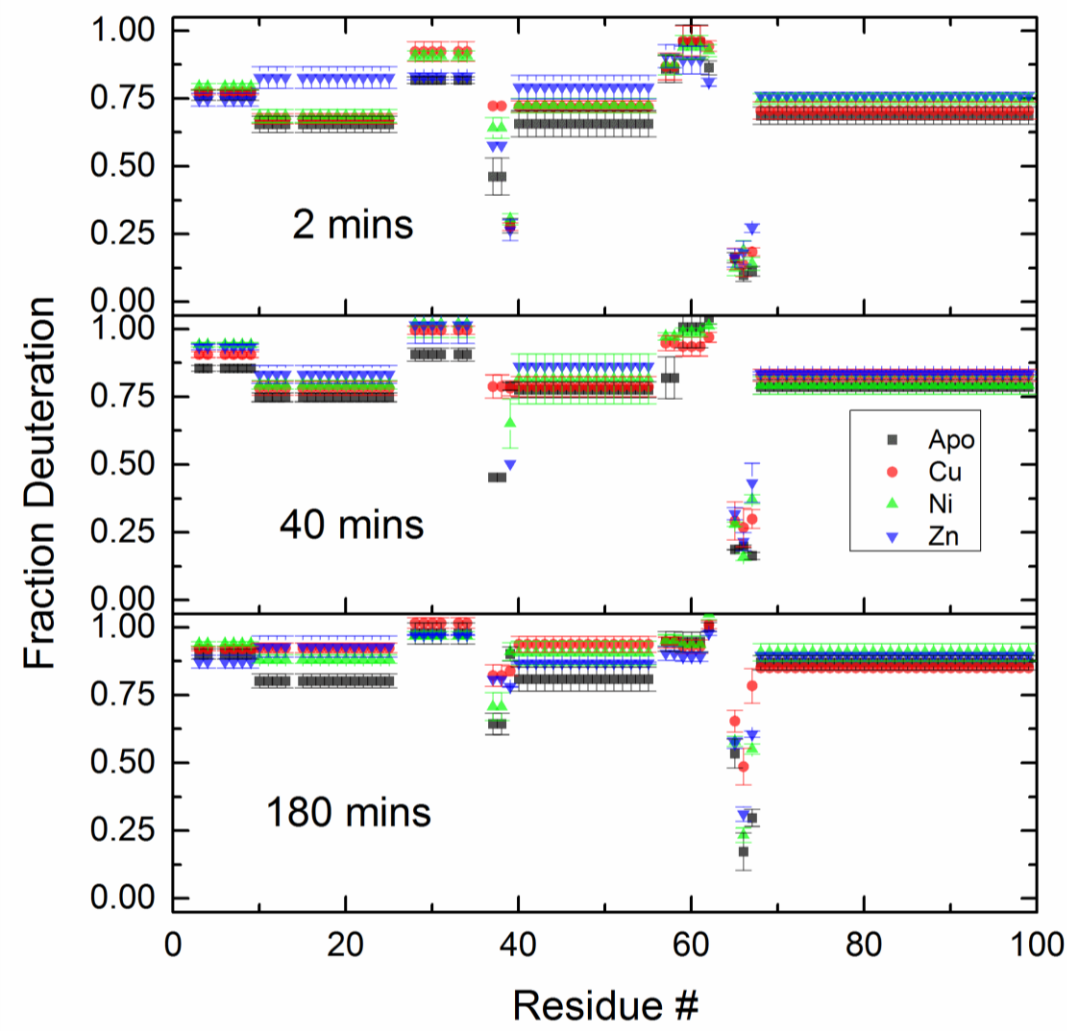

Figure 2-10 HDX uptake for each residue at after 2, 40, and 180 min of incubation in deuterium. These are derived from overlapping peptides produced by proteolysis.

Covalent labeling results with HNSB and DEPC indicate that Zn(II) causes β2m to adopt a very different structure than the Cu(II)-bound form of the protein. HNSB labeling of the Zn(II) bound protein and the metal free protein are essentially equivalent, indicating that Trp60 is not buried upon Zn binding (Figure 2-5). This result likely suggests that the *cis-trans* isomerization of the His31-Pro32 amide bond does not occur upon Zn binding. Moreover, DEPC labeling in the presence of Zn(II) leads to

significant deviations in labeling as compared to the Cu(II)-bound form, including at Lys19, His51, Lys58, Tyr67, Lys75, Ser88, and Lys94. Taken together, the HDX and covalent labeling data suggest that Zn(II)-binding does not result in the formation of the amyloidogenic conformer.

2.4 Discussion

The amyloid formation by β2m is thought to proceed by the formation of an amyloidogenic precursor that is formed by the *cis-trans* isomerization of the His31-Pro32 amide bond. This isomerization initiates the re-positioning of a number of residues including: Asp59, Arg3, Phe30, and Trp60.[50,60,53,112,62,63] The binding of Cu(II) to monomeric β2m has been shown to induce these structural changes and initiate oligomerization.[63,43] The ability of Cu binding to induce oligomerization and its binding site at a nearby residue (His31) lead to the thought that metal binding near Pro32 lowered the barrier to the *cis-trans* isomerization.[44,46] To further investigate the influence of transition metals on the structure of β2m, the complexes of Zn(II) and Ni(II) were investigated.

The binding of these three metals to β2m leads to significantly different outcomes. While the binding of Zn to β2m leads to aggregation, the aggregates are SDS-soluble amorphous aggregates, suggesting that Zn initiates a significantly different pathway than Cu. These amorphous aggregates are consistent with the fact that Zn binds to a very different site on the protein than Cu(II).[22] In contrast, Ni(II) binding does not stimulate any oligomerization or aggregation, as the protein remains monomeric,

despite the fact that Ni(II) binds to a similar site as Cu(II).[22] These different outcomes suggest that unlike Cu Zn and Ni do not promote the formation of the amyloidogenic precursor. To provide structural support for the hypothesis, we applied HDX and CL together with mass spectrometry. Most of β2m undergoes only minor structural changes upon the binding either metal. There are, however, regions that are affected differently by each metal. Zn binding initiates the most significant structural changes, leading to increased exchange in several locations even at the shortest time points, suggesting decreased stability. This decrease in stability between the D and E β strands and the A and B loop likely explains the propensity for the complex to amorphously aggregate. One region that undergoes increased HDX and that is particularly intriguing is the region near the B and E β strands. This region also demonstrates an increase in dynamics upon the binding of Cu. Because Zn also causes increased dynamics in this region but does not initiate amyloid formation, structural fluctuations at these sites are clearly not sufficient to induce the formation of the amyloid precursor.

Covalent labeling of Trp60 provided an important indicator of the formation of the amyloidogenic precursor. Only Cu(II) of the three metals was able to induce the burial of Trp60; burial of this residue is a key signature of the amyloidogenic precursor.[63,64] DEPC labeling also revealed other residues that undergo changes in solvent accessibility, and only Cu(II) binding show changes that are consistent with the amyloidogenic precursor as each metal has different effects on the solvent accessibility of residues including Lys19, His51, Lys58, Tyr67, Lys75, Ser88, and Lys94. The differences

in solvent accessibility upon binding each metal further suggest that each metal promotes the formation of different structures.

The use of both HDX and CL provides complementary information that can be used to generate a more detailed understanding of the changes in β2m structure upon metal binding. The use of HNSB, in particular, is interesting because it is an easy method for monitoring the formation of the amyloidogenic species of β2m. The ability to easily monitor this species could allow one to probe the rate at which the amyloidogenic precursor is formed under different conditions.

2.5 Conclusions

In this work we demonstrate that only Cu(II) binding to β2m, and not Ni(II) and Zn(II) binding, is capable of inducing its transformation into an amyloid competent state. The results described here highlight the fact that specific structural changes, and not just any partial unfolding, are essential for achieving the amyloidogenic state of this protein. This conclusion is most evident from a comparison of the structural changes caused by Cu(II) and Ni(II). While Ni(II) binds the protein in a similar manner as Cu(II), it appears to be unable to cause the *cis-trans* isomerization of the His31-Pro32 amide bond, which is essential for the formation of the amyloidogenic conformer. Zn(II) binding increases the overall structural dynamics of β2m, but these increased dynamics do not result in the formation of the amyloidogenic conformer. Instead, the overall decrease in stability caused by Zn(II) binding leads to the amorphous aggregation of β2m observed in the presence of this metal. In addition to obtaining some new insight

into the specific structural changes that are necessary for β2m amyloid formation, this work also demonstrates that covalent labeling of Trp60 by HNSB has the potential to measure the formation rate of the amyloidogenic conformer, which could be a useful tool for understanding more deeply the factors that influence β2m amyloid formation.

CHAPTER 3
INVESTIGATING THERAPEUTIC PROTEIN STRUCTURE WITH DIETHYLPYROCARBONATE LABELING AND MASS SPECTROMETRY

This chapter is adapted from a paper published as: Borotto, N.B., Zhou Y.,Hollingsworth, S. R., Hale J. E., Graban, E., Vaughan, R. C., Vachet, R. W. (2015) Investigating Therapeutic Protein Structure with Diethylpyrocarbonate Labeling and Mass Spectrometry. ***Anal. Chem****. (In press)*

3.1 Introduction

Protein therapeutics are the fastest growing segment of the pharmaceutical market, accounting for one-third of the overall late-stage drug development pipeline. They are anticipated to represent 20% of the total pharmaceuticals market value by 2017.[171] One key element in ensuring the safety and efficacy of these biologic drugs is the ability to measure and control the three dimensional (3D) structure of the protein active ingredients. In contrast to more traditional small molecule therapeutics, however, obtaining accurate, high resolution measurements of protein structures has proven to be extremely challenging.

Current structural techniques fall into two major categories: (1) rapid, low resolution techniques and (2) time and sample intensive, high resolution techniques.[85] Intrinsic fluorescence, circular dichroism (CD), dynamic light scattering (DLS), differential scanning calorimetry (DSC), and activity assays are examples of the first type. These methods provide an ensemble average of structures or are sometimes insensitive to

certain structural changes. NMR and X-ray crystallography are important examples of powerful high resolution techniques, but these methods are time-consuming, require a large amount of protein, and are not amenable to all proteins. Thus, there is a growing need for other techniques that can provide better resolution than the first category of techniques but do so in way that is easier and faster than the second category of techniques.

Mass spectrometry (MS)-based techniques offer an alternative because they can be rapid, provide moderate resolution, and can be sample efficient. Accordingly, these techniques have begun to fill an important niche in protein therapeutic analyses. The primary techniques used for monitoring protein solution structure by MS are hydrogen/deuterium exchange (HDX), chemical cross-linking, and covalent labeling. In HDX the mass spectrometer is used to measure the exchange of amide hydrogens for deuterium (or *vice versa*), and the extent of exchange at individual sites provides an indication of solvent accessibility and protein dynamics near that site. HDX/MS has been widely used to analyze protein structure[130,131,172–174] and recently has been applied to characterize the structure of protein therapeutics.[139,140,142,175] One challenge associated with HDX/MS is the transient nature of the label. As a result, special care and often expensive instrumentation are required to minimize back exchange and to accurately locate deuterated sites.

Methods that use covalent bond formation to characterize protein structure are not subject to back exchange. They also provide complementary information by reporting on protein side chains. Chemical cross-linking typically uses bifunctional

reagents to link residues that are spatially adjacent despite being distant in linear sequence. The cross-linked peptides are then sequenced and identified by MS, thereby revealing nearby residues. This method has been used to probe the structures of individual proteins[155] and protein complexes.[156,158–160] While this technique is not commonly used to study protein therapeutics, it has been used for antibody epitope mapping.[163] Other covalent labeling techniques use monofunctional reagents to monitor residue solvent accessibility as a means of probing structure. Hydroxyl radical footprinting (HRF) is the most common of these techniques.[148–151] In this method, hydroxyl radicals are produced through radiolysis or photolysis of water or hydrogen peroxide, and the resulting radicals then oxidize solvent accessible sites on the protein. Because of its broad reactivity and success with other protein systems, HRF has recently been applied to monitor structural changes in therapeutic proteins.[83,114] The technique was shown to be quite sensitive to subtle structural changes as it was able to distinguish expired protein therapeutics from fresh ones.[83] HRF also demonstrated the ability to identify the regions of aggregation in therapeutic monoclonal antibodies (mAbs).[114] While HRF shows great promise for studying therapeutic proteins, there are some challenges associated with implementation. Most notably, oxidation by hydroxyl radicals can produce over 50 different types of modifications, which can complicate MS analysis.[150] Moreover, in its most commonly used forms, a laser or synchrotron source is necessary to generate the radicals, which adds complexity and limits its wide applicability.

Another approach to covalent labeling uses amino acid-specific reagent molecules to modify solvent exposed residues. A wide range of reagents are available, ranging from those that have narrow specificity (e.g. succinimides) to those that have broad reactivity (e.g. DEPC).[164] This approach to labeling is simple as it requires no specialized equipment and typically produces only a single type of product, facilitating mass spectral analysis. While amino acid-specific reagents have been widely used to probe monomeric and oligomeric proteins,[164] their application to therapeutic proteins has been very limited. To our knowledge, only the carboxylate-specific reagent pair of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and glycine ethyl ester (GEE) has been used to probe the structure of a mAb.[74] This particular labeling chemistry is relatively simple to implement; however, because it is limited to only Asp and Glu residues, it results in relatively poor coverage of the protein's surface area and thus low effective structural resolution. Another reagent with broader reactivity such as DEPC should maintain the simplicity of this type of covalent labeling, while at the same time increasing resolution. Because DEPC is capable of labeling all nucleophilic residues, our group has shown that this reagent is capable of monitoring approximately 30% of surface residues of the average protein.[65,176] Such broad reactivity has enabled this reagent to provide insight into protein structure as well as protein-metal and protein-protein interactions.[54–56,65,164] In this work, we demonstrate the ability of DEPC labeling to assess structural perturbations in protein therapeutics by investigating three proteins before and after stressed conditions. We find that covalent labeling is capable of

identifying stress-induced structural perturbations in protein therapeutics, including the interface through which the protein therapeutics aggregate.

3.2 Experimental Procedure

3.2.1 Materials

Diethylpyrocarbonate (DEPC), imidazole, iodoacetamide, L-cysteine, papain from papaya latex, tris(2-carboxyethyl)phosphine (TCEP), and DL-dithiothreitol (DTT) were obtained from Sigma Aldrich (St. Louis, MO). The mAb immunoglobulin G1 (IgG1) was purchased from Waters Corporation (Milford, MA). Human β-2-microglobulin (β2m) was obtained from Fitzgerald Industries International (Concord, MA). Recombinant Human Growth Hormone (HGH) was purchased from Biovision (San Francisco, CA). Urea was purchased from Acros Organics (Geel, Belgium). Both immobilized trypsin and chymotrypsin were obtained from Princeton Separations (Adelphia, NJ). Sodium phosphate monobasic monohydrate was purchased from EM Science (Darmstadt, Germany). Sodium phosphate dibasic anhydrous, hydrogen peroxide, methanol, formic acid, acetonitrile, and water were purchased from Fisher Scientific (Fair Lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from Millipore (Burlington, MA).

3.2.2 Sample Preparation

β2m and IgG1 were prepared in 50 mM ammonium acetate (pH 7.4) and 50 mM sodium phosphate buffer (pH 7.4), respectively. Both proteins were then incubated at 75 °C for 15 min (IgG) or 1 day (β2m) for thermal degradation conditions. Oxidative conditions were carried out by incubating the protein in the presence of 3% H_2O_2 (w/w) at room temperature for 1 day. HGH was prepared in 10 mM sodium phosphate buffer (pH 8.0), and incubated at 65 °C for 2, 12, and 24 hours. After the forced degradation conditions, the proteins were reacted with DEPC and then analyzed by MS.

3.2.3 DEPC Labeling Reactions

Stock solutions of DEPC were prepared in acetonitrile. The DEPC reactions of β2m were performed for 1 min at 37 °C and were initiated by adding DEPC in a molar excess of 2.5. The total reaction volume for the experiments was 100 µL, and the total amount of acetonitrile added was 1%. Based on our previous work, this low percentage of acetonitrile has no noticeable influence on protein structure.[54–56,65] Experiments with β2m were performed in triplicate. The reactions were quenched after 1 min by adding 10 mM imidazole. Labeling of IgG1 (5 µM) was performed at a protein to DEPC molar ratio of 1:4 in a 50 mM phosphate buffer at pH 7.4. The solution was reacted for 5 min at room temperature before quenching by the addition of imidazole at a 1:50 (DEPC:imidazole) ratio. Five replicate reactions and analyses were conducted on the IgG1 samples. DEPC labeling of HGH was performed at a 1:5 (protein:DEPC) ratio for 1

min at room temperature. The amount of acetonitrile added was 1%. The reaction was quenched by the addition of imidazole at an 80 molar excess to DEPC. Three replicate reactions and analyses were conducted on the HGH samples.

3.2.4 Proteolytic Digestion

Since β2m and HGH have a single disulfide bond, TCEP (protein:TCEP=1:40 molar ratio) was added to reduce the disulfide bond and iodoacetamide (100 µM) was added simultaneously at room temperature for 30 min in the dark to alkylate the reduced Cys residues. The resulting samples were incubated with 10% (vol/vol) acetonitrile at 50 °C for 45 min prior to digestion with immobilized chymotrypsin (enzyme/substrate ratio of 1:10) for B2m and immobilized trypsin (1:10) for HGH at 37 °C. After 2 hours, the reaction mixture was centrifuged for 2 min at 9000 relative centrifugal force (RCF) to separate the enzyme from the protein. After that, the samples were either immediately analyzed by LC/MS or flash frozen in liquid nitrogen and stored at -80°C until LC/MS analysis.

To achieve complete digestion of IgG1, an initial digestion with activated papain was necessary. To activate papain 0.5 µM of the enzyme was incubated in 1 mM EDTA and 10 mM L-cysteine at 37 °C for 30 min. Once complete the cysteine concentration was reduced to less than 2 µM via four spins with a 3000 MWCO filter. The papain digestion was then performed for 2.5 hours using a 1:100 (papain:IgG1) ratio. IgG1 was then denatured and its disulfide bonds were reduced in a 50 mM phosphate buffered solution at pH 7.4 with 1 M urea and 20 mM DTT at 60 °C for 20 min. The reduced

disulfides were then alkylated with 40 mM iodoacetamide for 2 min. Immobilized trypsin was then added to achieve a 1:3 (enzyme:substrate) ratio, and the digestion reaction was allowed to proceed overnight at room temperature. After digestion, the samples were spun at 9200 RCF for 5 min, and the supernatant was collected, flash-frozen in liquid nitrogen, and stored at -80 °C until analysis via LC/MS.

3.2.5 HPLC Separation

β2m HPLC separations were conducted using an HP1100 HPLC system (Agilent, Wilmington, DE) with a Discovery C18 column (15 cm × 2.1 mm, 5 µm particle size; Supelco, St. Louis, MO, USA). Peptide fragments from the proteolytic digests were eluted using a linear gradient of methanol containing 0.1% acetic acid that increased from 10% to 100% methanol over 30 min at a flow rate of 0.25 mL/min.

HPLC separations of IgG were performed using a Thermo Scientific Acclaim PepMap RSLC C18 (15 cm x 50 µm, 2 µm particle size) on an Easy-nLC 1000 system (Thermo Scientific, Tewksbury, MA). To achieve sufficient separation of the proteolytic peptides, a shallow gradient was used where %B (0.1% formic acid in acetonitrile) was increased from 0 to 40% over 90 min. The column was then flushed by jumping to 95 % B over 15 min. It was then held at 95% B for the remainder of the separation (i.e. another 20 min). A flow rate of 0.225 µL/min was used.

HPLC analyses of HGH were performed using an Accela LC system (Thermo Scientific, Tewksbury, MA) with a ZORBAX 300SB-C18 MicroBore RR column (1.0 X 150 mm, 3.5 µm particle size, Agilent, Wilmington, DE). Peptides were eluted over a 50-

minute gradient where %B was increased from 2% to 45% over the first 35 minutes, then elevated to 80% for an additional 5 minutes. The column was then re-equilibrated with 0% B for 10 minutes. Blank runs were run in between each sample, and were each run at a flow rate of 0.05 mL/min.

3.2.6 Mass Spectrometry

On-line high performance liquid chromatography (HPLC) MS analyses were performed on all protein digests. The HPLC details can be found in the SI. Mass analysis of β2m proteolytic fragments was carried out on a Bruker AmaZon (Billerica, MA, USA) quadrupole ion trap mass spectrometer equipped with an electrospray ionization source. Typically, the electrospray needle voltage was kept at ~4 kV, and the capillary temperature was set to 250 °C. Either collision-induced dissociation (CID) or electron transfer dissociation (ETD) was used to obtain tandem mass spectra.

For IgG, MS analyses were performed using a Thermo Scientific Orbitrap Fusion (Tewksbury, MA) mass spectrometer. The electrospray ionization source was typically operated at a needle voltage of 2200 volts, and the ion transfer tube temp was set to 300 °C. Tandem mass spectra were collected using CID with a normalized collision energy of 35%. Due to the large number of measured peaks, an exclusion limit of 60 sec was activated after five spectra were collected for any given peak. The resolution of the Orbitrap was set to 60000.

MS analysis for HGH was performed using a Thermo Scientific LTQ-XL Orbitrap (Tewksbury, MA) mass spectrometer equipped with an electrospray ionization source.

The ESI needle voltage was kept at 5 kV. Tandem mass spectra were generated using an HCD collision energy of 35.

3.2.7 Peptide Identification

Raw mass spectral data files were converted to .mgf format using msconvert software.[177] The .mgf files were analyzed with SearchGUI.[178] The search engines X!tandem,[179,180] MS Amanda,[181] MS-GF+,[182] OMSSA,[183] and Comet[184] were all used. Spectra were searched against a database constructed from the cRAP database (http://www.thegpm.org/crap/index.html) with the sequence of the proteins of interest added. Spectra were searched against the custom database and against the reverse, decoy database. Variable modification by DEPC of the residues H, Y, K, T, S and the protein N-terminus was added as a user modification (mass addition of 72.0211). Variable oxidation of M was also used in searches. Carboxyamidomethylation of cysteine was used as a fixed modification. Unspecific enzyme cleavage was selected, and a precursor mass tolerance of 10 ppm was used.

Search data were visualized using Peptideshaker[185] with protein, peptide and PSM FDRs set at 1%. PTMs were scored using the PhosphoRS algorithm. Identification features were exported in .csv format, and these features were used to construct a custom database for peak identification in MZmine.

3.2.8 Peptide Peak Quantification

Raw data files were imported into MZmine,[186] and mass detection was done in centroid mode at the MS1 level. Chromatograms were constructed, deconvoluted and peak identification was performed with a custom database constructed from the Peptideshaker export. When multiple files were analyzed, deconvoluted spectra were aligned using the RANSAC algorithm. The quantified, identified and aligned data were exported to a .csv file using the export function.

3.2.9 Circular Dichroism

All IgG solutions were diluted to 0.75 μM in 50 mM phosphate buffer at pH 7.4 prior to analysis. For HGH conditions all solutions were diluted to 5 μM in 10 mM phosphate buffer at pH 8.0. Circular dichroism was measured using a J-715 spectropolarimeter (Jasco, Easton, MD). The scan ranged from 250 to 195 nm with a scan resolution of 0.5 nm, a scan rate of 100 nm/min, and a response time of 1 sec. Raw data were converted into mean residue ellipticity using the CD Analysis & Plotting Tool (CAPITO).[187]

3.2.10 Fluorescence

The concentration of IgG was diluted to 0.75 μM in 50 mM phosphate buffer at pH 7.4 prior to analysis. For HGH, the protein was diluted to 5 μM in 10 mM phosphate buffer at pH 8.0. A Photon Technology International Quantamaster-4SE (PTI, Edison, NJ)

was used to obtain fluorescence spectra for IgG. Tryptophan fluorescence was collected using an excitation wavelength of 295 nm and slit widths of 1 nm. Emission scans ranged from 310-440 nm. A Synergy H1 multi-mode plate reader (BioTek, Winooski, VT) was used to measure the fluorescence of HGH. Samples were excited at 295 nm and emissions were monitored from 300-440 nm.

3.2.11 Dynamic Light Scattering

A Zetasizer Nano-ZS (Malvern Instruments, Worcestershire, U.K.) was used to measure the hydrodynamic radii of native and heat denatured IgG. A 1 mL solution of 1 µM IgG was used for these experiments. Five runs were conducted for each sample, and volume particle size distribution is reported. For HGH the measurements were carried out using a Zetasizer Nano-S (Malvern Instruments, Worcestershire, U.K.). HGH was measured at 5 µM after treatment by the conditions indicated. Measurement duration was according to preset levels, and intensity/volume distributions of the samples from at least four runs were recorded in each dataset.

3.2.12 Size Exclusion Chromatography

For SEC experiments, the protein was separated using a SuperSW2000 30 cm × 4.6 mm column (GE Healthcare Biosciences, Piscataway, NJ) installed on an Agilent HP1100 series HPLC system (Wilmington, DE). Before injection of the sample, the SEC column was first equilibrated with a 20 mM ammonium acetate mobile phase (pH = 7.4)

at a 0.035 mL/min flow rate for 1 h. 20 µL of the protein sample was injected into the sample loop. A variable wavelength detector set at 214 nm was used for detection.

3.3 Results and Discussion

3.3.1 β-2 microglobulin (β2m)

β2m shares the same β-sandwich fold as each of the domains in IgG and thus was chosen as an initial model system. The protein under native and both heat- and oxidatively-degraded conditions was probed using DEPC. Labeling was done at a 2.5:1 molar ratio (DEPC:protein) as previous work from our group had demonstrated that a labeling ratio of 4:1 or less provided good labeling yield without significantly perturbing a protein's structure.[65]

The modification results for the β2m residues under all three conditions reveal that β2m undergoes noticeable structural changes upon exposure to elevated temperature and H_2O_2 (Table 3-1). The residues that are labeled under native conditions are displayed as spheres in Figure 3-1A and B. All 16 of these residues are found on the exterior of the protein and are exposed to solvent, which is consistent with previous DEPC labeling results for this protein under native conditions.

Table 3-1. Modification percentages for individual residues of β2m before and after heating at 75 °C for 24 hours or oxidation with 3% H2O2 for 24 hours.

B2m	Native (control)			Heated for 1 day			Significant*	3% H_2O_2			Significant*
N-term	49	±	1	44	±	2	yes	7.6	±	0.3	yes
Lys6	2.5	±	0.1	3	±	0.5	no	0	±	0	yes
Ser11	20	±	3	11.8	±	0.3	yes	6	±	2	yes
His13	32	±	2	17	±	1	yes	4	±	1	yes
Lys19	2.7	±	0.4	2	±	2	no	0	±	0	yes
Ser20	1.1	±	0.3	4.8	±	0.7	yes	1.7	±	0.3	no
Ser28	0.25	±	0.05	0	±	0	yes	0.3	±	0.09	yes
His31	1.9	±	0.2	0.72	±	0.08	yes	0	±	0.1	yes
Ser33	0.68	±	0.09	0	±	0	yes	0.18	±	0.01	yes
Lys41	0.6	±	0.1	1	±	0.1	yes	0	±	0	yes
Lys48	0	±	0	1.2	±	0.2	yes	0	±	0	no
His51	2.2	±	0.3	0	±	0	yes	0.27	±	0.07	yes
Ser55	1.4	±	0.2	0	±	0	yes	0	±	0	yes
Ser57/ Lys58	2.5	±	0.4	0	±	0	yes	0	±	0	yes
Tyr67/ Thr68	2.2	±	0.3	2	±	1	no	6	±	2	yes
Lys91	2.1	±	0.4	2	±	0.2	no	0.1	±	0.2	yes
Lys94	3.1	±	0.6	3	±	0.4	no	0.2	±	0.3	yes

*A difference was considered significant if the p-value, calculated by performing an unpaired T-test, was less than 0.05 (corresponding to a 95% confidence level, n=3)).

After being exposed to thermal stress for 24 hours, the extent and pattern of labeling significantly changes. The residues highlighted in blue in Figure 3-1A undergo statistically significant decreases (p-value < 0.05) in labeling extent, while the residues in red undergo an increase in labeling extent. Because many residues undergo a decrease in labeling extent and these residues are clustered on one face of the protein, these results suggest that the protein aggregates upon overnight heating. Support for this conclusion is found from SEC measurements (Figure 3-2), which reveal protein aggregates are formed. The residues that decrease in labeling extent (i.e. Ile1, Ser28, His31, Ser33, Ser55, and Ser57) are likely at the interface(s) of these oligomers. Moreover, the same region of the protein is buried in the MHC complex that β2m forms

physiologically.[28] This region's exposure is likely thermodynamically unfavorable in the monomeric protein, potentially explaining its propensity for aggregation at this site. Interestingly, two residues (i.e. Lys41 and Lys48) on the loop that connects two of the aggregating β strands show an increase in labeling after heating, suggesting that this region of the protein unfolds upon heating to facilitate aggregation.

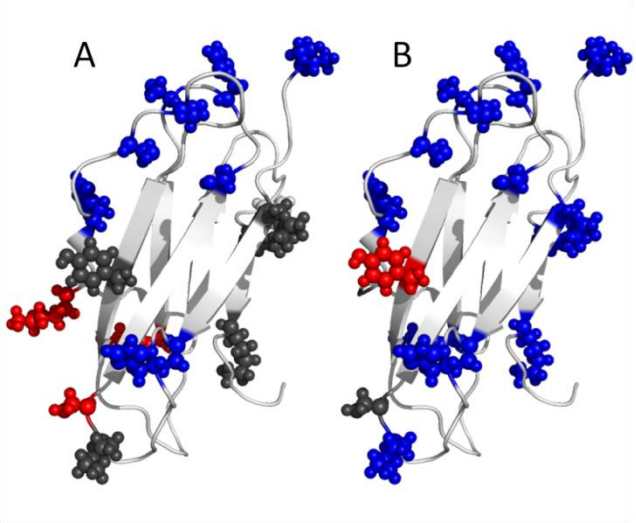

Figure 3-1 Covalent labeling results for β2m. Spheres represent residues that were labeled with DEPC. The color indicates whether the residue has undergone any significant change in labeling after being exposed to a perturbing condition (blue: decrease, red: increase, gray: no change). A) Heating at 75°C for 24 hours. B) Oxidation with 3% H2O2 for 24 hours. Changes in covalent labeling are mapped onto the NMR structure of β2m (PDB accession code: 2XKS).

Covalent labeling of the oxidized protein also indicates that the protein aggregates after exposure to H_2O_2. Nearly all residues undergo a decrease in labeling extent (Figure 3-1B), signifying extensive aggregation. Analysis via SEC corroborates this conclusion (Figure 3-2). While many of the same residues decrease in labeling as in the heated sample, the additional residues indicate the aggregates are also mediated by other interfaces. Overall, the covalent labeling experiments with β2m successfully

demonstrate the ability of this technique to identify structural changes, especially interfacial sites that are formed upon heat- and oxidatively-induced aggregation.

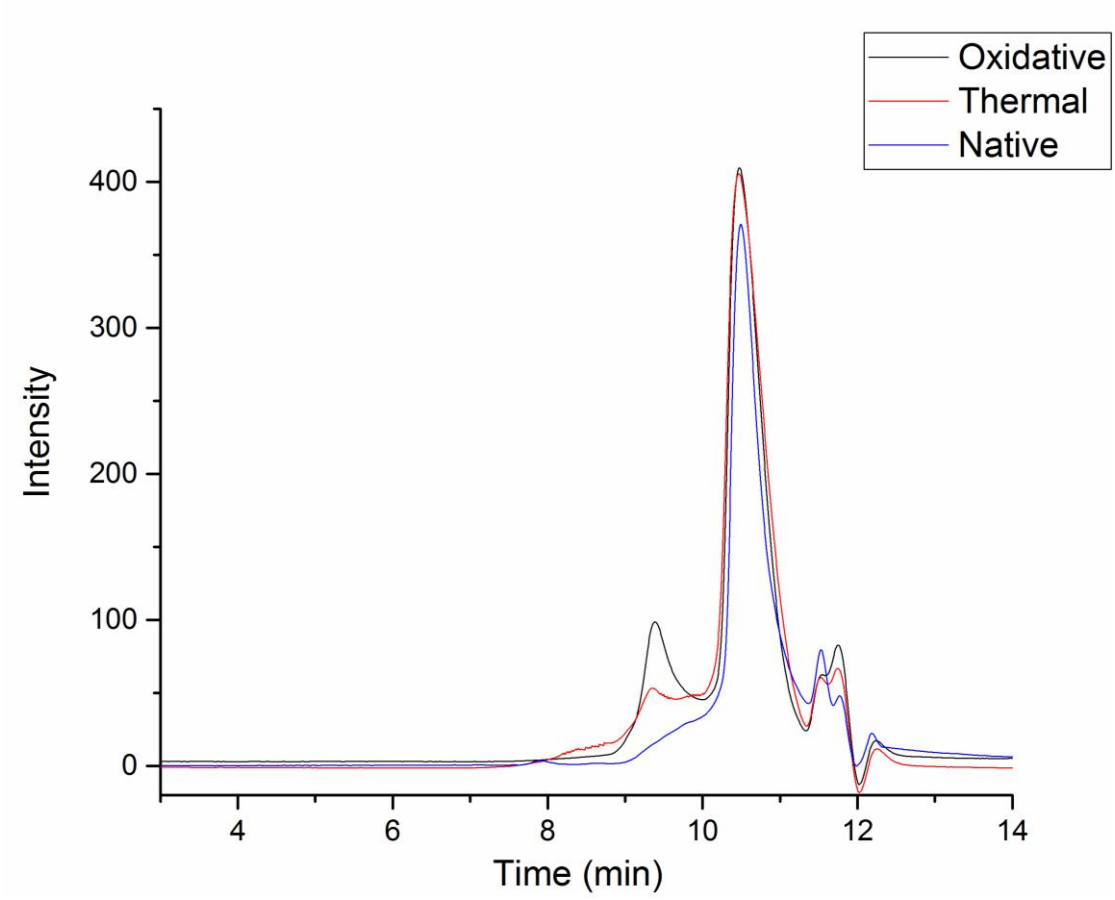

Figure 3-2 Size exclusion chromatography of β2m before (blue) and after heating (red) and oxidation (black). Both chromatograms demonstrate the presence of aggregated species. These aggregated complexes are evident from the peaks eluting earlier than 10.5 min.

3.3.2 Immunoglobulin G (IgG)

The promising results with β2m prompted us to apply the method to IgG under native and thermally degraded conditions. To minimize structural perturbations to IgG during the labeling reaction, we limited the DEPC:protein ratio to 4:1. We also monitored the protein's structure using CD and fluorescence spectroscopy. Both techniques demonstrate that the protein undergoes no significant structural

perturbations after reacting with DEPC at these concentrations (Figure 3-3), confirming the structural integrity of the protein. We were pleased to find that DEPC labeling of IgG under these conditions results in the labeling of almost 30% of the amino acids in the protein (Table S1 and S2 in the SI).

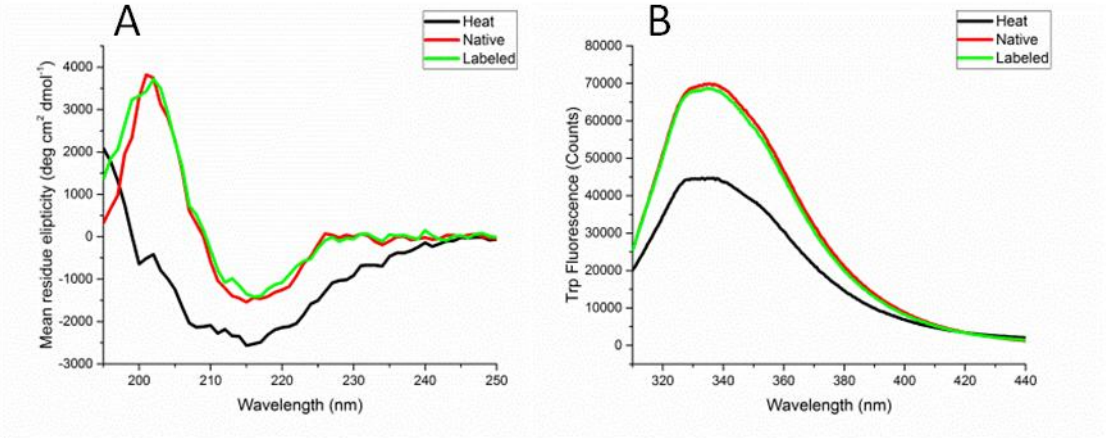

Figure 3-3 Circular dichroism (A) and tryptophan fluorescence (B) of IgG1 under normal (red), heated (Black), and DEPC-labeled (green) conditions. The essentially identical overlap between the spectra of the normal and DEPC-labeled samples demonstrates that covalent labeling has little effect on the structure of IgG1.

Upon comparing the labeling results of the thermally degraded sample with those generated under native conditions, we find that numerous sites undergo changes in labeling. Figure 3-4 summarizes these results by showing the percent change in DEPC labeling that each residue undergoes upon heat treatment relative to the unheated sample. Figure 3-4A illustrates the labeling for IgG's light chain, while Figure 3-4B shows the results for the heavy chain.

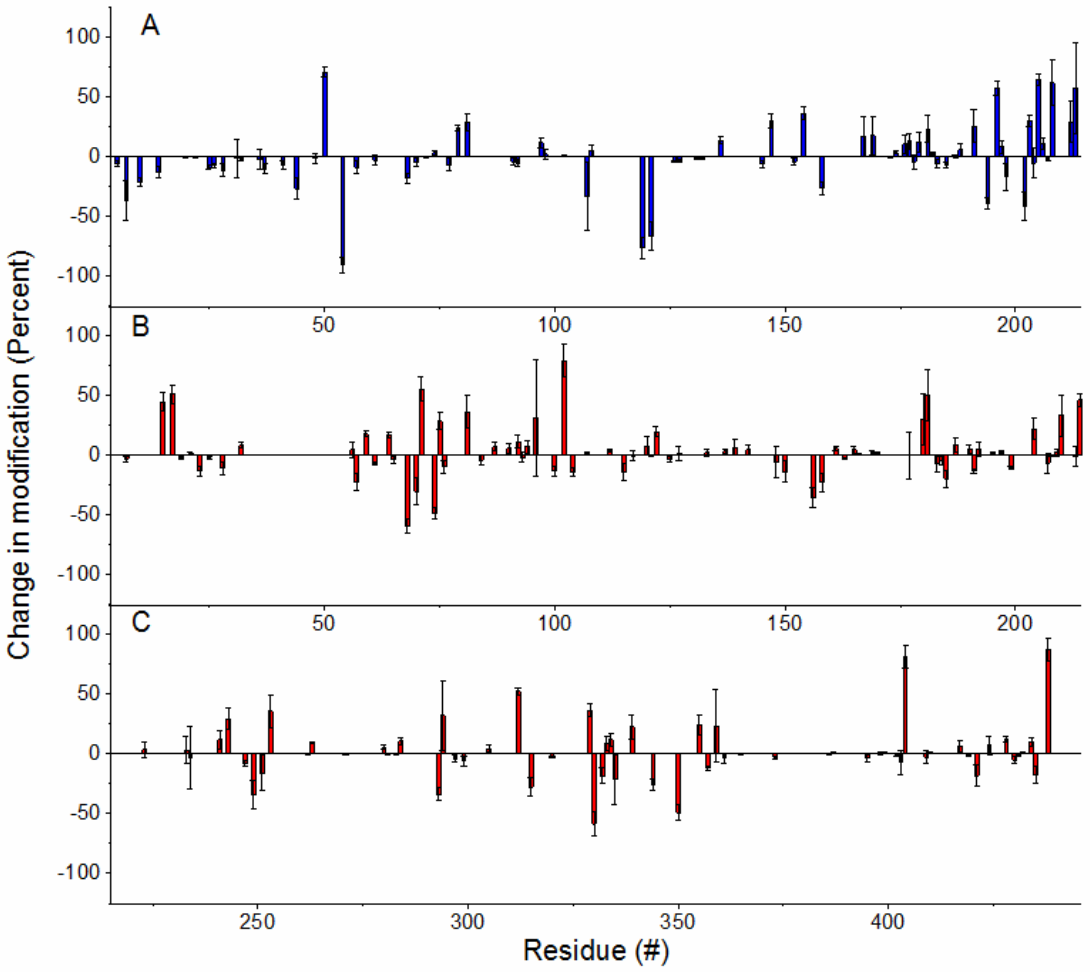

Figure 3-4 Bars represent changes in modification of IgG1 after heating from experiments involving five replicates. Negative values represent residues that are more protected after heating. A) Light chain. B) Top: V_H and C_H^1 domain of heavy chain. Bottom: C_H^2 and C_H^3 domain of heavy chain

Because about 200 residues are labeled in the protein, these results were further simplified by considering only the statistically significant changes in each domain. The significant relative changes were broken into bins based on the magnitude of the change (Table 3-2). Each domain has a relatively equal number of residues undergoing increases or decreases in labeling; however, almost all of the residues in the light chain's variable domain (V_L) undergo significant decreases in labeling. This clustering of residues

suggests that the protein might be aggregating upon heating, and this domain mediates this aggregation. Indeed, DLS demonstrates that IgG undergoes significant aggregation after 15 min of heating at 75 °C (Figure 3-5).

Table 3-2. Number of residues within each domain of IgG1 whose relative labeling change after heating is statistically significant and whose value falls within the labeled bin.

Domain[a]	< -80%	< -40%	< -10%	Total Decrease	> 200%	> 100%	> 10%	Total Increase
V_L	12	8	1	21	4	0	0	4
C_L	10	3	3	16	12	3	2	17
V_H	13	2	1	16	13	1	1	15
C_H^1	9	2	1	12	13	4	2	19
C_H^2	9	5	0	14	4	3	3	10
C_H^3	5	6	0	11	4	1	3	8

aCL and VL represent the constant and variable domains of the light chain. VH and CH1-3 represent the variable and the three constant domains of the heavy chain, respectively.

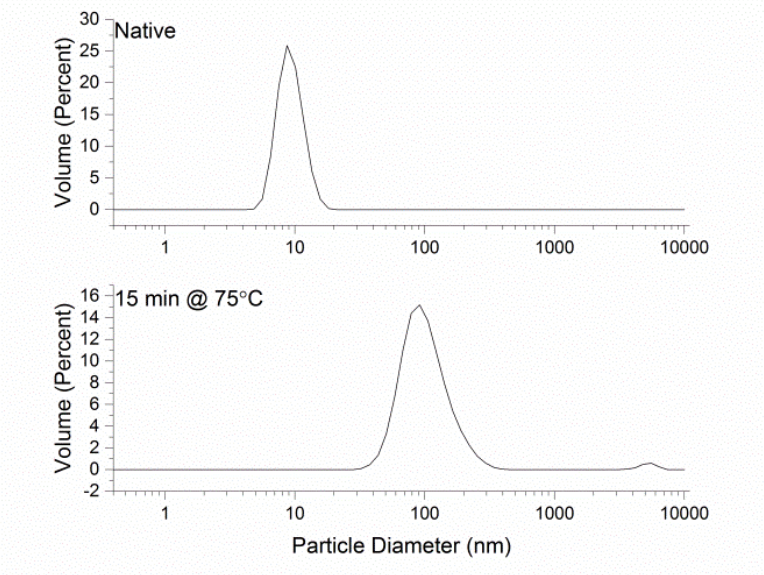

Figure 3-5 Dynamic light scattering data for IgG1 before (top) and after heating at 75 °C for 15 minutes (bottom). These data demonstrate that IgG1 aggregates upon heating.

Greater structural insight is obtained by mapping the data in Table 3-2 onto an IgG homology model. A homology model was generated by the Swiss-Model workspace[188,189] using an IgG crystal structure (PDB: 1IGY) as a template. As expected, the V_L domain presents a large cluster of residues that undergo a decrease in labeling (Figure 3-6). Somewhat surprisingly, mapping the labeling data also reveals another potential interface on the V_H domain, which also shows a clustering of residues undergoing a decrease in labeling. A color coded depiction of the different IgG domains in a space-filling model can be found in Figure 1-3.

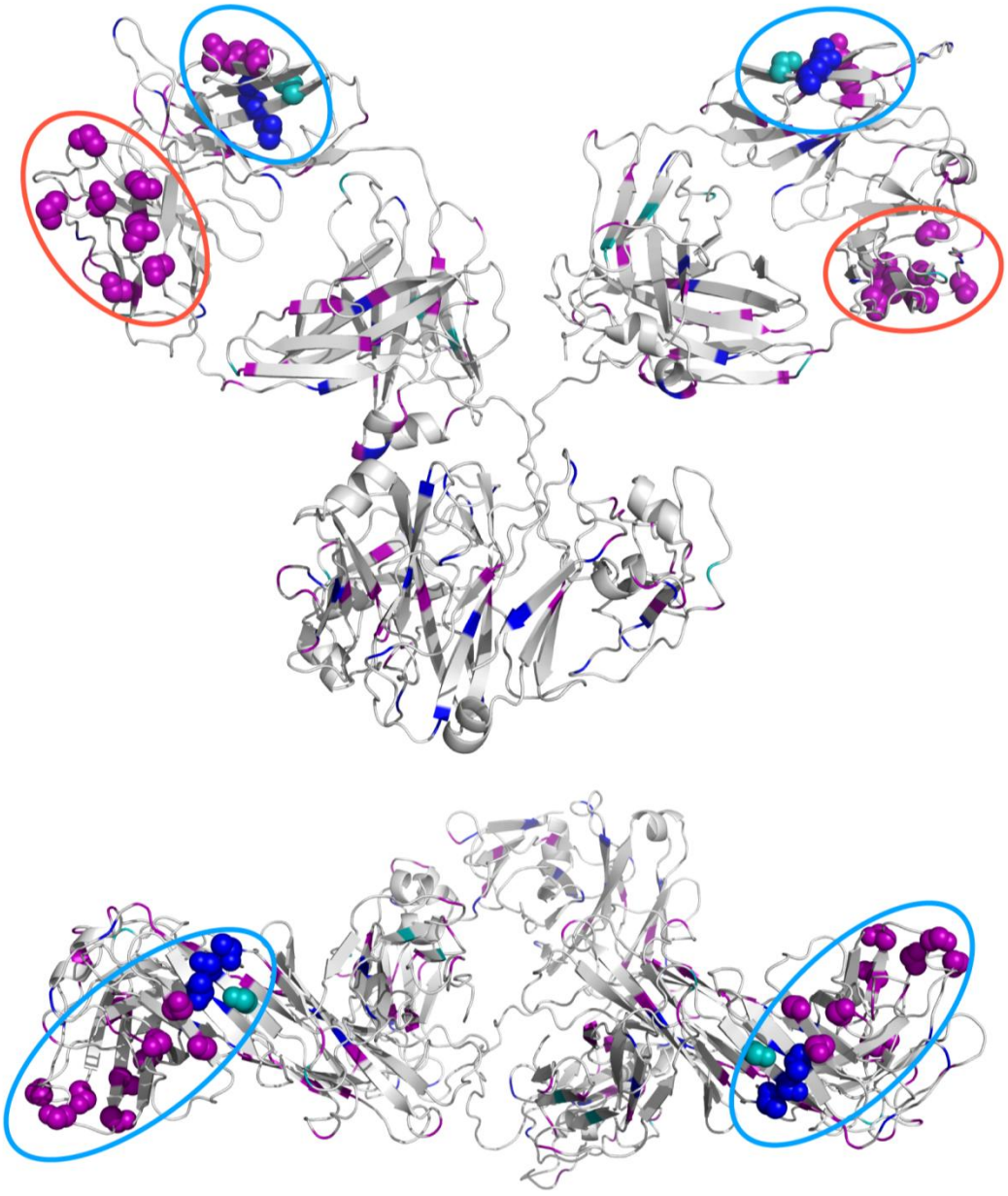

Figure 3-6 Cartoon representations of IgG1 homology model. Side view (top) and top view (bottom). Spheres represent residues that are likely at the aggregate interface. Colors represent the magnitude of the reduction (Purple: >80%, Blue: 40-80%, and Teal: 10-40% reduction in labeling). The likely interfaces on the V_L and V_H domains are circled (V_L: blue and V_H: red).

The possible role that the V_L and V_H domains play in mediating IgG1 aggregates is further supported by the aggregation predictor tool Zyggregator (Figure 3-7).[190–193] When Zyggregator prediction data are overlaid onto IgG1's structure (Figure 3-7), there

are four surface accessible regions in the protein that have a cluster of residues with a strong propensity to aggregate. Two of the four predicted sites are the V_L and V_H domains, which are implicated to be involved in aggregation by covalent labeling (Figure 3-7). The other two predicted regions clearly have smaller surface areas, decreasing their likelihood of being true interfaces in the aggregates. It is interesting to note that previous studies are divided on which domains mediate aggregation processes in IgG's. Some work suggests that typically the C_H^2 domain is the primary site of aggregation,[85,194–196] while others suggest the variable regions in the Fab domain as the sites of aggregation.[89,114,197,198] It is quite possible that the aggregation site varies from antibody to antibody.[198] We feel our labeling results provide strong evidence for the role of the variable regions in the light and heavy chains in mediating aggregation in IgG1.

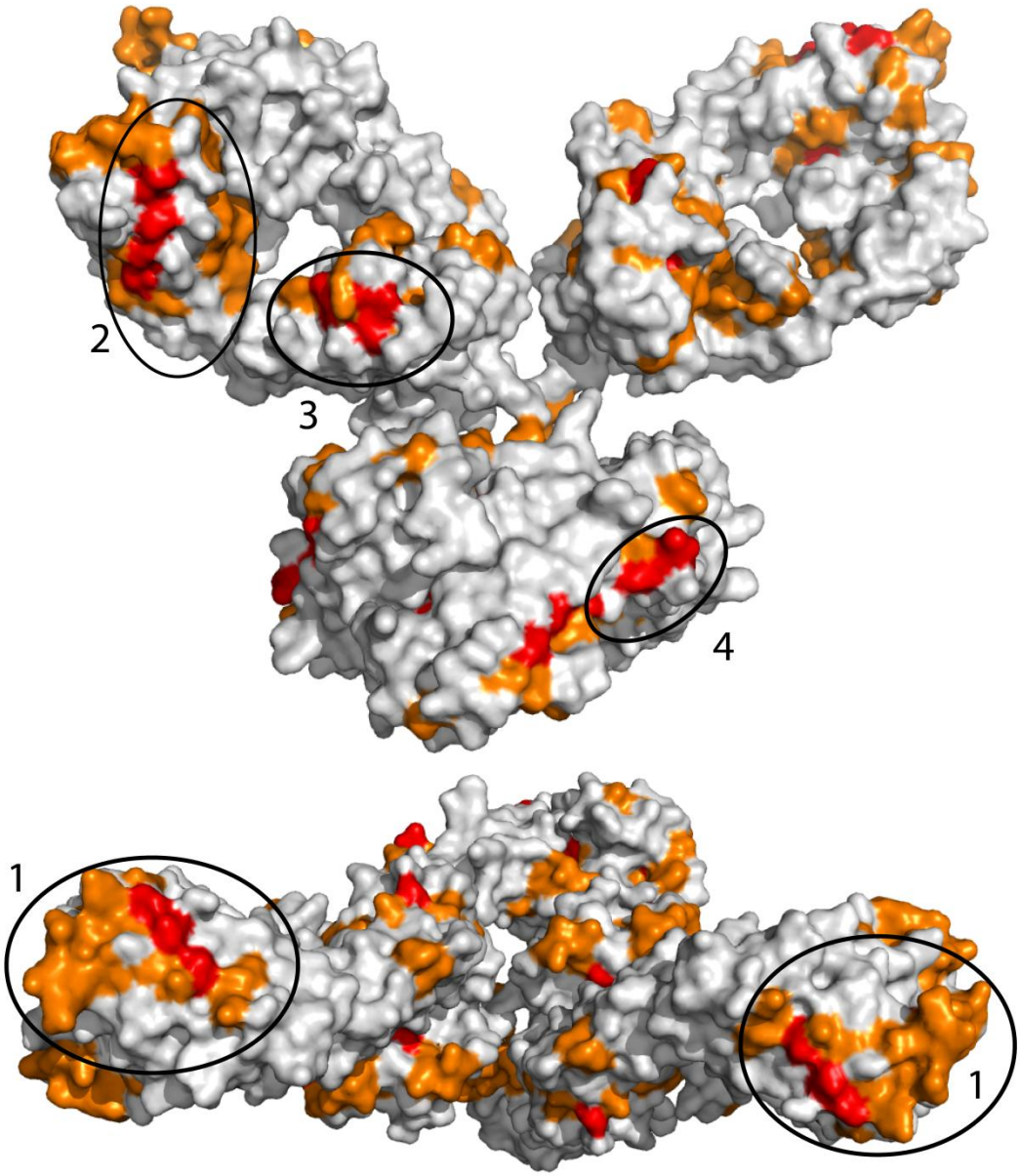

Figure 3-7 Propensities for aggregation from Zyggregator as indicated by Z-scores. Z-Score > 2 (red) Z-score > 1 (orange). A higher score denotes an increased likelihood of aggregation. Circled sites highlight the regions where there is a clustering of residues with a high calculated aggregation propensity. Labeled sites 1 and 2 are the covalent labeling indicated aggregation sites on the V_L domain and the V_H domain, respectively.

3.3.3 Human Growth Hormone

The results from IgG and β2m demonstrate that covalent labeling can identify proteins which have undergone severe structural perturbations. In order to test the technique's ability to identify minor structural changes, we studied HGH. Labeling of this protein was performed before and after heating at 65 °C, a temperature that is 12 °C below its melting temperature.[199] CD and intrinsic fluorescence spectra (Figure 3-8) both show only minor structural perturbations after heating to this temperature.

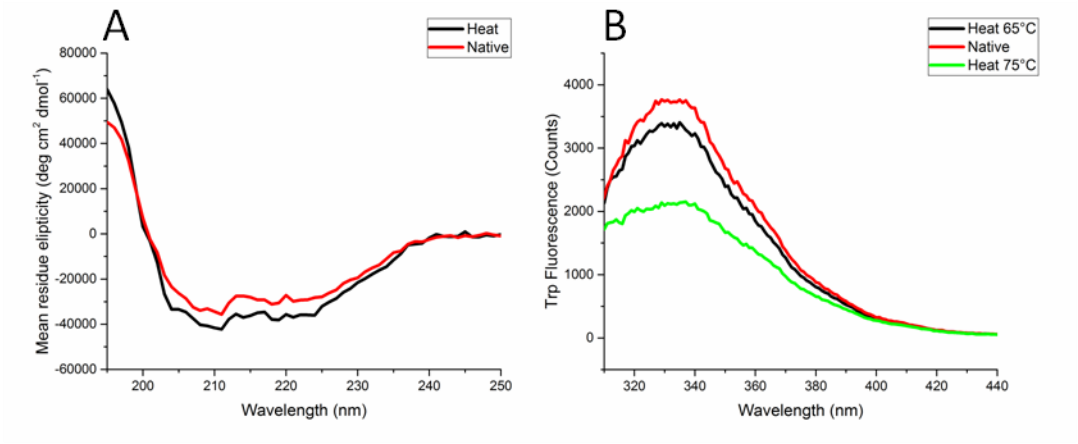

Figure 3-8 Circular dichroism (A) and tryptophan fluorescence (B) of HGH under normal (red), heated to 65 ºC (black), and heated to 75 ºC (green) conditions. Both fluorescence and circular dichroism show only minor changes in structure after heating. The fluorescence spectrum at 75 ºC is included as a positive control to demonstrate that fluorescence can reveal structural perturbations for HGH.

Upon DEPC labeling, 41 modification sites are identified (Table S5 in the SI) corresponding to over 20% of the protein. This amount of labeling ensures sufficient coverage of the protein's structure. The labeling percentages of all the labeled sites for both native and heat-denatured HGH are summarized in Figure 3-9. When comparing the heat-denatured protein to the natively-structured protein, only six residues are found to undergo a significant change (p-value < 0.05) in labeling extent. They are His19,

Thr28, Thr61, Thr136, Lys159, and Tyr165. Of these six, two residues undergo increased labeling upon heating (Thr136 and Lys159).

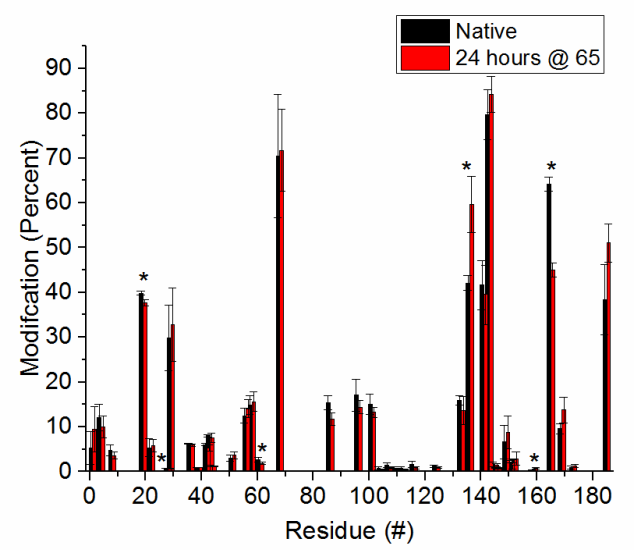

Figure 3-9 Covalent labeling results for HGH before and after heating at 65°C for 24 hours. Asterisks (*) indicate residues that have undergone a statistically significant change. A difference was considered significant if the p-value, calculated by performing an unpaired T-test, was less than 0.05 (corresponding to a 95% confidence level at n=3).

To understand the structural implications of these changes, we mapped the six residues on to a crystal structure of HGH (Figure 3-10, PDB: 1HGU). The two residues (Thr136 and Lys159) that undergo increased labeling after heating are on opposite ends of a long disordered region, signifying further melting of this region. The four residues (His19, Thr28, Thr61, and Tyr165) that undergo decreased labeling are clustered on the opposite face of the protein.

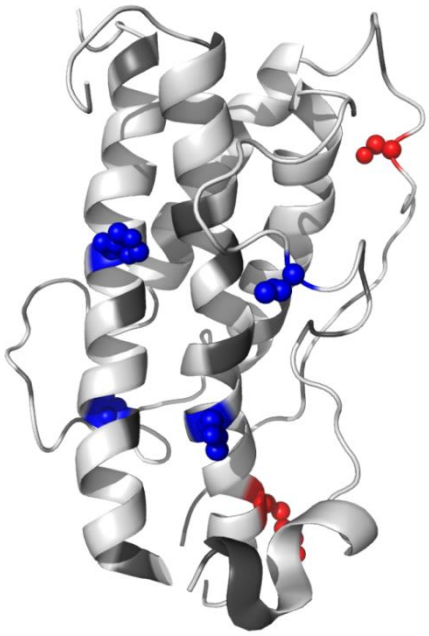

Figure 3-10 Summary of the covalent labeling results for HGH. Spheres represent residues that underwent significant changes in DEPC labeling after heating at 65 ºC for 24 h (blue: decreased, red: increased). Changes in covalent labeling are mapped onto a crystal structure of HGH (PDB accession code: 1HGU).

While such clustering might suggest aggregation, DLS measurements (Figure 3-11) indicate that the protein does not aggregate under these conditions. Instead, the DLS measurements reveal that the protein undergoes a slight compaction upon heating. Therefore, it is possible that the four residues undergo a decrease in labeling extent because they become less solvent exposed during this compaction process.

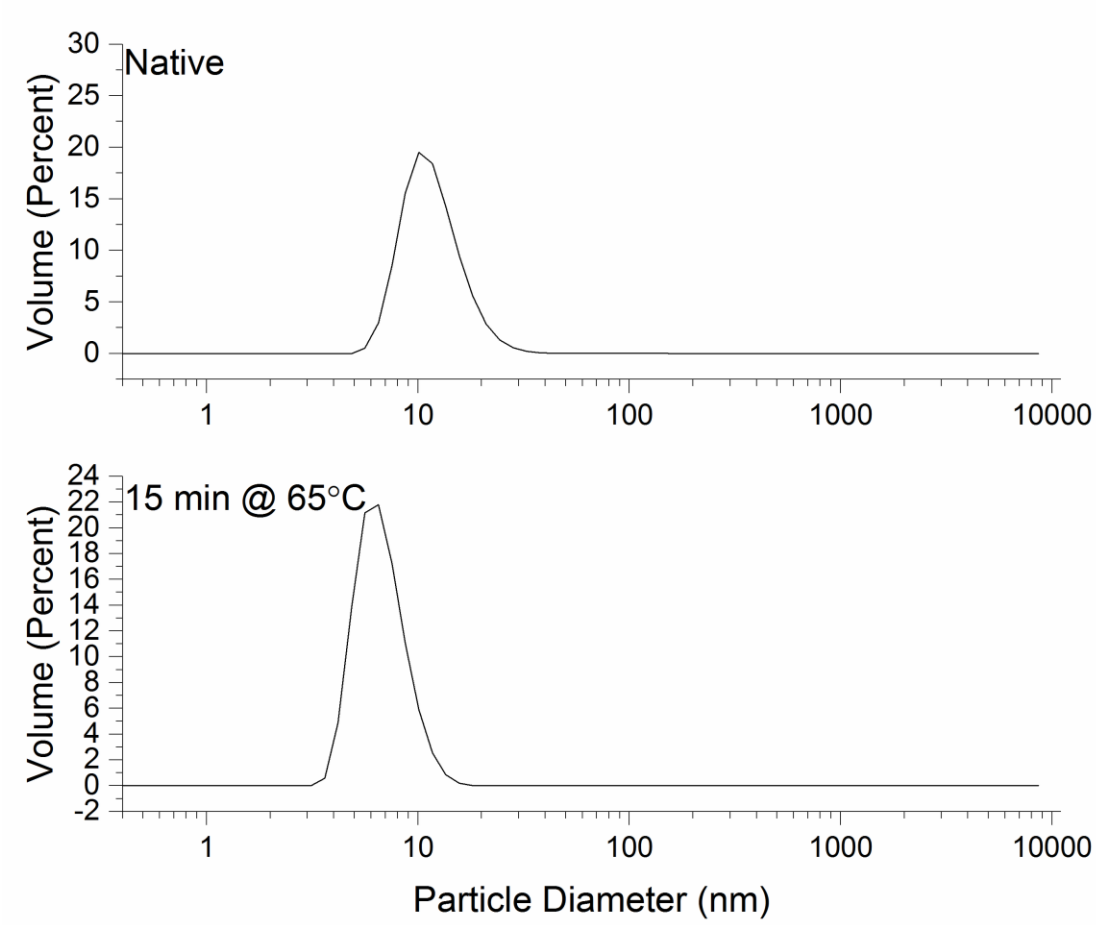

Figure 3-11 Dynamic light scattering data for HGH before (top) and after heating at 65 °C for 24 h (bottom). These data demonstrate that HGH undergoes some degree of compaction upon heating.

Under denaturing conditions it has been demonstrated that HGH maintains a majority of its helical structure. Its loops, however, are known to become significantly more dynamic than the rest of the protein's structure.[96] Repositioning of these loops might cause Thr136 and Lys159 to become more solvent exposed, while at the same time causing His19, Thr28, Thr61, and Tyr165 to become less solvent exposed. Overall, these data for HGH suggest that DEPC labeling with MS detection is sensitive enough to detect minor structural perturbations.

3.4 Conclusions

We have used DEPC-based covalent labeling as a means to monitor the structure of therapeutic proteins. Through the study of three proteins, β2m, IgG, and HGH, we have shown that DEPC labeling is capable of identifying specific structural perturbations that occur upon exposing these proteins to common forced-degradation conditions. Because the label can probe up to 30% of the residues in a protein, this method provides a high degree of structural resolution relative to other covalent labeling reagents. DEPC labeling is particularly valuable for identifying interfacial residues in protein aggregates. For example, this technique was able to identify the variable domains of the light and heavy chain as the regions that mediate aggregation of IgG1 upon heating. DEPC labeling is also able to distinguish relatively minor perturbations in protein structure as illustrated by the experiments with HGH. Given the high effective resolution provided by DEPC labeling and the ease with which it can be performed relative to other MS-based techniques, we predict that this approach will be a powerful tool for studying therapeutic proteins.

CHAPTER 4
LABEL SCRAMBLING DURING CID OF COVALENTLY LABELED PEPTIDE IONS

This chapter is adapted from a paper published as: Borotto, N. B., Degraan-Weber, N., Zhou, Y., and Vachet, R. W. (2014) Label Scrambling During CID of Covalently Labeled Peptide Ions. ***J. Am. Soc. Mass Spectrom***. 25, 1739-46.

4.1 Introduction

Covalent labeling along with mass spectrometry is being increasingly used to study higher order protein structure and protein-protein complexes.[148–151,164,200–203] Covalent labels that are typically used either label specific amino acid residues (e.g. succinimides for lysines) or a range of amino acid residues (e.g. hydroxyl radicals). We have recently shown that diethylpyrocarbonate (DEPC) is a promising reagent molecule because of its ability to modify a wide range of amino acids.[55,56,65,164,165] This potential to modify numerous amino acids enables DEPC to probe approximately 30% of the average protein.[165] Usually, proteins that are labeled with DEPC are then subjected to proteolysis so that the modification sites can be pinpointed to individual residues, thereby improving the resolution of this method.

While DEPC has been successfully used to study the structures of proteins and protein complexes, this labeling reagent introduces an electrophilic site into the side chains of the residues it modifies, opening up the possibility for some unwanted chemistry. For example, we recently demonstrated that, in solution, cysteine residues have the ability to capture carbethoxy groups from other residues that were modified by DEPC. This unwanted label transfer can be eliminated by deactivating cysteine's

strong nucleophilic character via alkylation just after DEPC labeling of the protein is finished.[176] The presence of a new electrophilic site might also affect the gas-phase chemistry of DEPC-labeled peptides especially when subjected to slow collisional activation in a quadrupole ion trap mass spectrometer. Indeed, Reid and co-workers and others have reported that phosphorylated peptide ions can have their phosphate group transferred from one amino acid to another upon CID.[204–206] The result is the incorrect assignment of the phosphorylation site on the peptide. In addition to the scrambling of phosphate groups in peptide ions, methyl,[207] acetyl, and formyl groups[208] have also been documented to undergo transfer from one amino acid to another during CID in the gas phase.

With the known chemistry of DEPC and these previous studies in mind, we set out to investigate whether DEPC label scrambling can occur during CID of DEPC labeled peptide ions. Through the study of numerous peptides, we find that label transfer can occur in DEPC-labeled peptides, and it occurs with similar characteristics to phosphate group transfer in phosphorylated peptides.

4.2 Experimental Methods

4.2.1 Materials

Diethylpyrocarbonate (DEPC), imidazole, iodoacetamide, and tris(2-carboxyethyl)phosphine (TCEP) were obtained from Sigma Aldrich (St. Louis, MO). Angiotensin 1-10 (DRVYIHPFHL) , angiotensin 1-13 (DRVYIHPFHLVIH), apelin 13

(QRPRLSHKGPMPA), β amyloid (YEVHHQKLVFF), semastatin (SPWTKCSATCGGGHYMRTR), neuromedin-C (GNHWAVGHLM) and adrenocorticotropic hormone (ACTH) 1-13 (SYSMEHFRWGKPV) were obtained from the American Peptide Company (Sunnyvale, CA). Human β-2-microglobulin (β2m) was obtained from Lee Biosolutions (St. Louis, MO). Immobilized chymotrypsin and triethylamine acetate (pH 8.0) were obtained from Princeton Separations (Adelphia, NJ). Ammonium acetate, methanol, formic acid, acetonitrile, and water were purchased from Fisher Scientific (Fair lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from Millipore (Burlington, MA).

4.2.2 DEPC Labeling Reactions

Peptide solutions with a concentration of 100 μM in 10 mM ammonium acetate were reacted with DEPC at either a 1:1 or 1:4 (peptide:DEPC) ratio for 2-5 min at 37°C. A 6 mM stock solution of DEPC in acetonitrile was added to obtain the final DEPC concentration. The final DEPC concentration resulted in a total acetonitrile concentration that was less than 2% (v/v) of the reaction mixture. The reaction time and ratio were chosen to ensure that the unmodified and the singly modified peaks were the most prominent products. The protein β2m was reacted with DEPC at a 1:4 (protein:DEPC) ratio for 1 min at 37°C. In all cases, the DEPC reactions were quenched by the addition of 10 mM imidazole.[65]

4.2.3 Proteolytic Digestion

Prior to proteolytic digestion of β2m, the quenching agent, imidazole, was removed using a 10,000 MWCO filter, resulting in a 40 µL solution of 250 µM β2m. β2m was then incubated in a buffer solution (100 mM triethylamine at pH 8.0) with 10% (v/v) acetonitrile at 50 °C for 45 min. Next β2m was reacted with TCEP (1:40 ratio) and iodoacetamide (1:80 ratio) in the dark for 30 min to reduce its disulfide bond and alkylate the resulting free thiols. Immobilized chymotrypsin was then added to achieve a 1:10 ratio of enzyme to substrate. The digestion reaction was allowed to proceed at 37 °C for 2 hours.

4.2.4 HPLC Separation

HPLC separations were performed using a Supelco Discovery C18 column (15 cm x 2.1 mm, 5 µm particle size, St. Louis, MO) on an HP1100 HPLC system (Agilent, Wilmington, DE). To achieve sufficient separation of DEPC-labeled peptide isomers, an isocratic elution was used (40% acetonitrile in water with 0.1% formic acid) for 20 min. This was followed by a 0.5%/min increase of acetonitrile for an additional 20 min. A flow rate of 0.35 mL/min was used, and the effluent from the column was split by four prior to introduction into the mass spectrometer. The proteolytic digests were eluted using a linear gradient of methanol ranging from 10 – 100% methanol over 30 min. This mobile phase contained 0.1% acetic acid and had a flow rate of 0.25 mL/min.

4.2.5 Mass Spectrometry

Mass analysis was performed using either a Bruker Esquire-LC or a Bruker AmaZon (Billerica,MA) quadrupole ion trap mass spectrometer. These instruments are equipped with electrospray ionization sources, and the needle voltage was typically set to 3000 V. The capillary temperature was set to 300°C. Both CID and ETD were used to obtain tandem mass spectra. The ion isolation width for both methods was set to 1.0 Da. CID voltages were typically between 0.65 and 1.1 V and were chosen to achieve optimal dissociation efficiency. For ETD experiments, the low m/z cutoff was typically set to 135, and the reaction time was typically set to 150 ms.

4.3 Results and Discussion

4.3.1 CID for the Label Site Identification

CID is commonly used to determine label locations in covalently labeled peptides and is typically successful. As an example, CID of the peptide apelin 13 (QRPRLSHKGPMPA) can readily enable the identification of the DEPC-labeled residues upon a 2 minute reaction with this reagent (Figure 4-1). LC/MS analysis of the labeled peptide suggests that there are three labeled sites as indicated by the chromatogram in Figure 4-1a. By examining the sequence ions upon CID of each peak in the chromatogram, the specific amino acid residues labeled by DEPC can be assigned. For DEPC-labeled apelin 13, the CID spectrum of the second chromatographic peak shows a

series of b ions that are all labeled, and several y ions that are not labeled, indicating that the N-terminus is the labeled site (Figure 4-1b). The CID spectrum of the third chromatographic peak is consistent with a label on Lys8, as suggested by labeled y_6, y_7, y_9, and y_{11} product ions, unlabeled y_4 and y_5 ions, labeled b_8 and b_{11} ions, and unlabeled b_2 and b_7 ions (Figure 4-1c). The final chromatographic peak corresponds to a labeled His7 product. This assignment is evident from the labeled y_7, y_9, and y_{11} product ions, unlabeled y_4, y_5, and y_6 ions, labeled b_7, b_8 and b_{11} ions, and an unlabeled b_2 ion (Figure 4-1d).

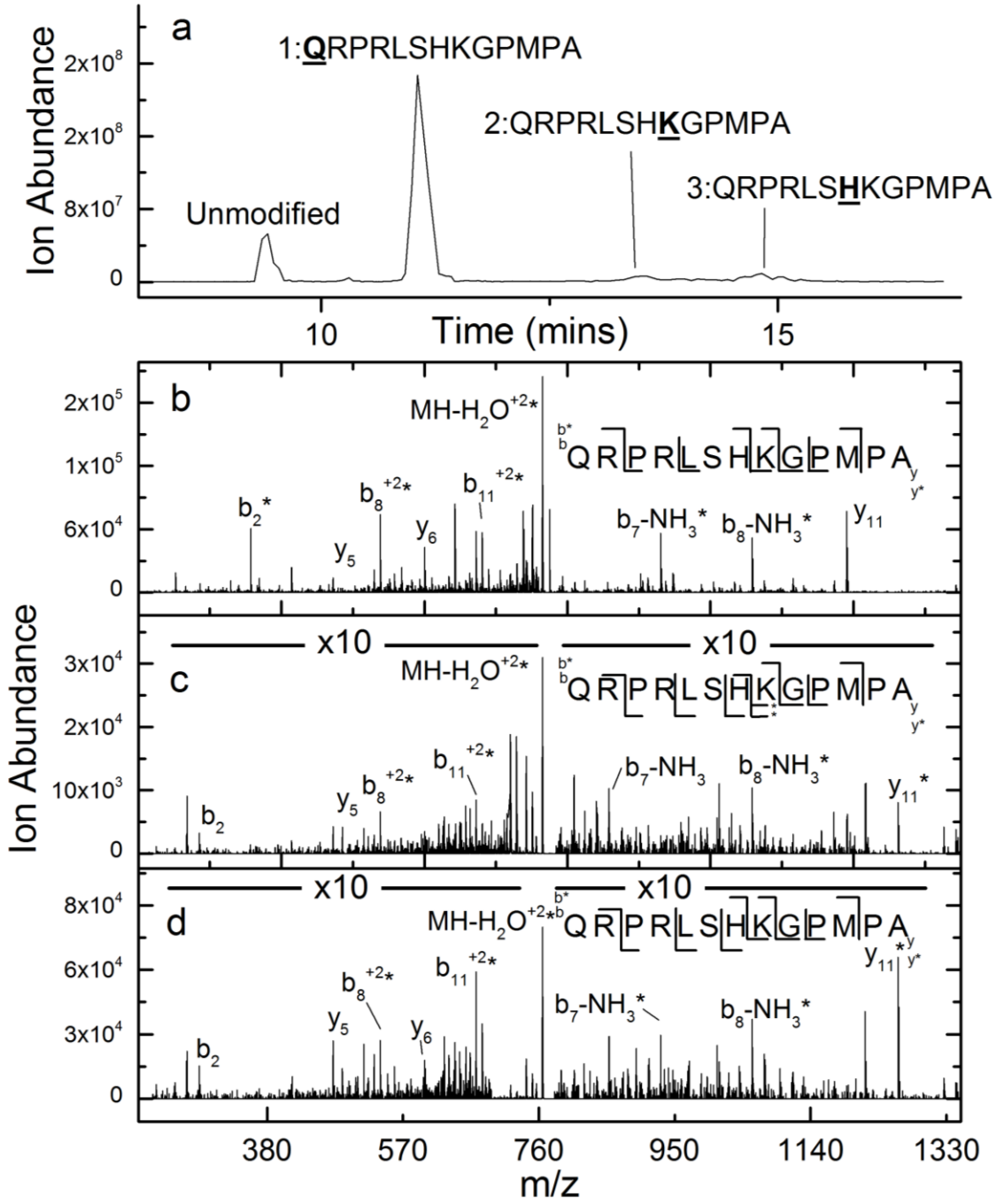

Figure 4-1 (a) Total ion chromatogram after HPLC separation of DEPC-labeled apelin 13. (b) CID spectrum of the $(M+2H)^{2+}$ ion of the second chromatographic peak from the DEPC-labeled sample of apelin 13, indicating that the N-terminus is modified. (c) CID spectrum of the $(M+2H)^{2+}$ ion of the third chromatographic peak from the DEPC-labeled sample of apelin 13, indicating that Lys8 is modified. *Interfering ions make assignment of the labeled and unlabeled versions of the y_6 ion somewhat ambiguous. (d) CID spectrum of the $(M+2H)^{2+}$ ion of the fourth chromatographic peak from the DEPC-labeled sample of apelin 13, indicating that His7 is modified.

4.3.2 Ambiguous Label Assignment by CID

While CID often provides correct labeling site information, as exemplified by the data for apelin 13, we have also found evidence that covalent labels can be scrambled from one site to another during the CID process. The CID data of DEPC labeled-angiotensin I illustrates this phenomenon. Upon labeling this peptide with DEPC for 5 min, on average a single modification site per peptide is measured, resulting in four isomers as indicated by the chromatogram in Figure 4-2a. Each peak in the chromatogram corresponds to a chemically distinct structure, and indeed the CID data for the first (Figure 4-2b) and second (Figure S1 in the Supplemental Information) labeled peaks are consistent with this assertion, as Tyr4 and the N-terminus, respectively, are identified as the modified sites.

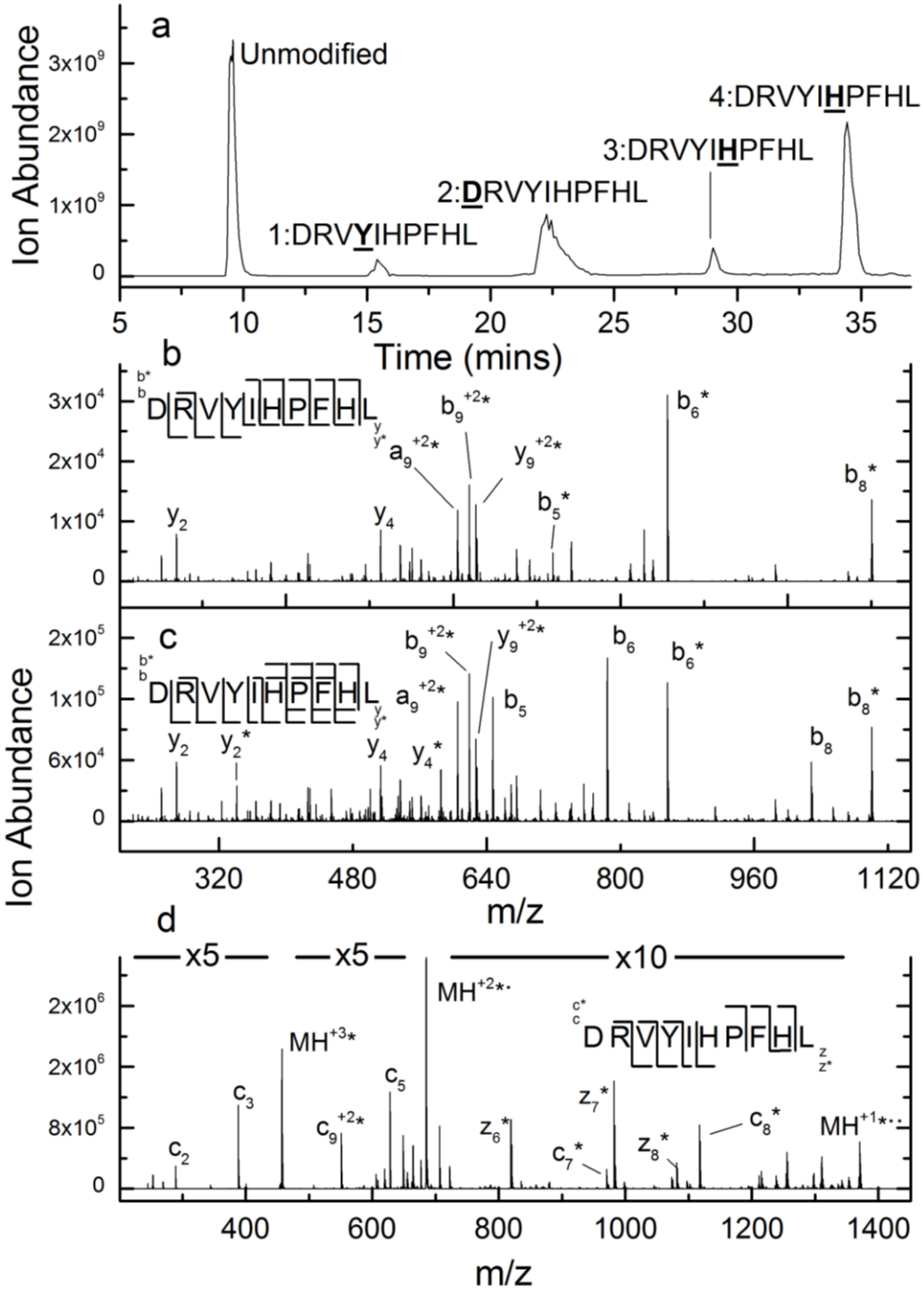

Figure 4-2 (a) Total ion chromatogram after HPLC separation of DEPC-labeled angiotensin. (b) Example CID spectrum of the $(M+2H)^{2+}$ ion of the second chromatographic peak from the DEPC-labeled sample of angiotensin I, indicating that Tyr4 is modified. Unlabeled y_2 to y_6 ions, labeled y_7 to y_9 ions, labeled b_5 to b_9 ions, and an unlabeled b_2 ion indicate that Tyr4 is labeled. (c) CID spectrum of the $(M+2H)^{2+}$ ion of the final chromatographic peak from the DEPC-labeled sample of angiotensin I. The presence of both labeled and unlabeled b6, b8, y4, and y2 product ions causes the labeled site to be ambiguous. (d) ETD spectrum of the $(M+3H)^{3+}$ ion of the final chromatographic peak from the DEPC-labeled sample of angiotensin I, indicating that His6 is labeled.

The identity of the labeled sites for the modified peptides associated with the final two chromatographic peaks, however, are not as clear. This fact is most evident from the presence of both labeled and unlabeled versions of the b_6, b_8, y_4, and y_2 product ions in both spectra (Figure 4-2c and S2). One possible explanation for this is that these chromatographic peaks correspond to two unseparated isomers – one modified at His6 and the other at His9. This possibility is unlikely considering the great care taken to identify LC conditions to separate any isomers present. Moreover, the ETD spectra of these peaks (Figure 4-2d and S3) indicate that only His6 is modified with no evidence for the modification of His9. This conclusion comes from the series of c and z product ions in which ions that include His6 are labeled and all that do not are unlabeled. Incidentally, the ETD spectra of peaks 1 and 2 lead to the same conclusions about modification sites as the CID data (see Figures S4 and S5). Another possible explanation is that the DEPC label is lost as a neutral. This would, for example, explain the presence of both labeled and unlabeled b_6 and b_8 product ions in the CID spectrum; however, the absence of an unlabeled b_9 ion and no evidence for label loss during the CID of other DEPC-labeled peptides (e.g. apelin 13 in Figure 4-1) suggests that label loss is not the cause. Moreover, label loss cannot explain the presence of labeled y_2 and y_4 ions. These latter ions can only arise from the label being transferred to His9.

Upon considering the data for DEPC-labeled angiotensin I, we propose that the CID data for the final two chromatographic peaks of modified angiotensin I arise via a collision-induced transfer of the DEPC label from one His residue to another. Such a gas-

phase rearrangement of a side chain modification has been seen before[204–208] during CID analyses, as indicated in the introduction. The migration of the DEPC label from His6 to His9 during collisional activation of angiotensin I can explain the existence of all observed ions and is consistent with differences in dissociation mechanisms between ETD and CID. Unlike label loss, transfer of DEPC from His6 to His9 explains, for example, why the b_9 product ion is only observed as a labeled species. The fact that label migration does not take place during ETD makes sense due to the non-ergodic nature of its dissociation mechanism. We suspect that the scrambling of the label is driven by the fact that addition of DEPC to a histidine side chain results in an electrophilic site that is vulnerable to attack by other nucleophilic sites on the peptide. In the case of angiotensin, this nucleophilic site could be the other, unlabeled histidine residue. A possible mechanism is illustrated in Figure 4-3.

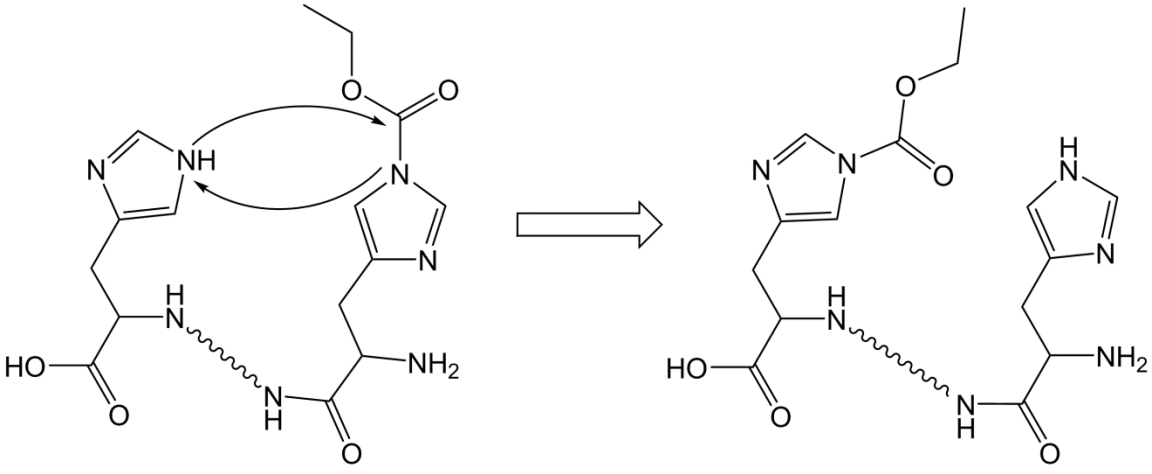

Figure 4-3 Possible mechanism of rearrangement, where the DEPC label on one histidine side chain is transferred to another via a nucleophilic attack.

Another peptide that appears to undergo label scrambling during CID is the peptide adrenocorticotropic hormone (ACTH). LC/MS analysis of DEPC labeled ACTH

indicates that two amino acids are labeled in this peptide (Figure 4-4a). In this case, the label sites associated with both chromatographic peaks from the labeled peptide are ambiguous when CID is used. The tandem mass spectrum of the first labeled peak suggests scrambling between the N-terminus, His6, and Lys11 (Figure S6) as indicated by the presence of labeled and unlabeled versions of the b_2 - b_{10} series of product ions and the y_6 - y_{11} series of product ions. The fact that the labeling site is clearly identified as the N-terminus when dissociated by ETD (Figure S7) suggests that unseparated isomers are not the source of this ambiguity. CID of the second labeled chromatographic peak suggests both His6 and Lys11 as the labeled sites based on the existence of labeled and unlabeled versions of the b_6 - b_{10} series of product ions and y_3 - y_7 series of product ions (Figure 4-4b). In contrast, the ETD spectrum of this labeled peptide readily identifies His6 as the labeled site (Figure 4-4c).

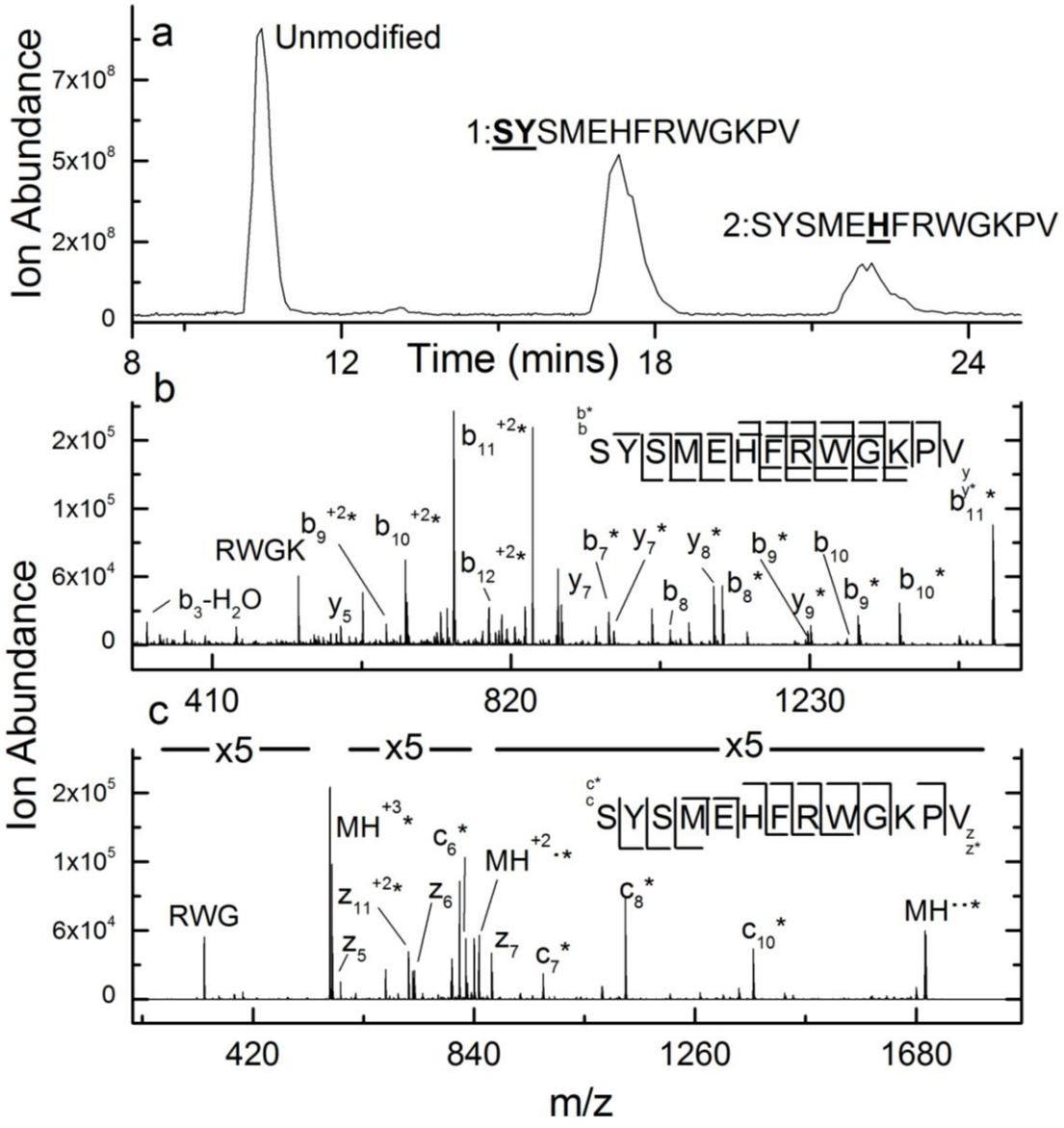

Figure 4-4 (a) Total ion chromatogram after HPLC separation of DEPC-labeled ACTH. (b) CID spectrum of the (M+2H)2+ ion of the first labeled chromatographic peak from the DEPC-labeled sample of ACTH. The presence of both labeled and unlabeled b10, b9, b8, and y7 product ions suggests label scrambling. (c) ETD spectrum of the (M+3H)3+ ion of the second labeled chromatographic peak from the DEPC-labeled sample of ACTH, indicating that His6 is labeled.

4.3.3 Effect of Charge State on Labeling Scrambling

Previous work has shown that the gas-phase rearrangement of phosphate and methyl groups in peptide ions is dependent on the precursor ion's charge state and the

subsequent proton mobility within the ion.[204,207] Phosphorylated and methylated peptide ions with non-mobile and partially mobile protons were found to have a higher propensity to rearrange. Proton mobility in this context refers to the number of net protons on the peptide ion as compared to the number of basic residues in the peptide. Non-mobile protons are present when the number of net protons is lower than the number of basic residues in the peptide. To test whether proton mobility influences DEPC label scrambling, we compared the CID spectra of the $[M+2H+DEPC]^{2+}$ and $[M+3H+DEPC]^{3+}$ ions of the fifth chromatographic peak of labeled angiotensin (i.e. peaks in Figure 4-2a). Interestingly, the different charge states result in drastically different dissociation patterns with regard to label scrambling. There are no longer any labeled y_2 or y_4 product ions observed upon dissociation of the +3 charge state (Figure 4-5a); these ions were key indicators of label scrambling for the +2 ion (see Figure 2c). In addition, the unlabeled b_6 and b_8 ions, which also implied label scrambling in the +2 ion, are substantially reduced in relative abundance.

A similar influence of proton mobility on label scrambling is also observed for ACTH. As shown earlier in Figure 4-4b, the +2 charge state of third chromatographic peak exhibits scrambling upon activation with CID. When the +3 charge state of this same labeled peptide is dissociated using CID, no evidence for scrambling is observed (Figure 4-5b). The absence of unlabeled b_6, labeled y_3, and y_7 product ions in this spectrum removes all ambiguity about the label assignment. The dramatic influence of charge state on the dissociation behavior of these DEPC-modified peptides further supports the notion that label scrambling is occurring.

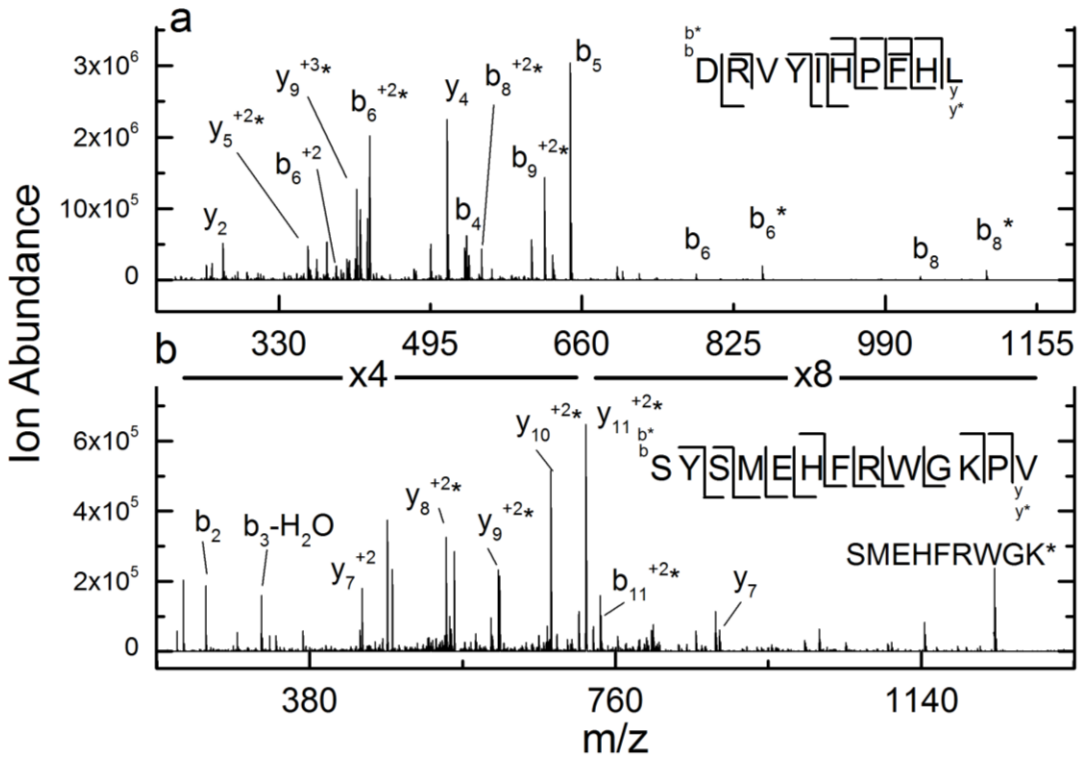

Figure 4-5 (a) CID spectrum of the $(M+3H)^{3+}$ ion of the final labeled chromatographic peak from the DEPC-labeled sample of angiotensin I. The absence of labeled y_2 and y_4 product ions indicates label scrambling is not occurring. (b) CID spectrum of the $(M+3H)^{3+}$ ion of the second labeled chromatographic peak from the DEPC-labeled sample of ACTH. The absence of unlabeled b_6, labeled y_3, and y_7 product ions in this spectrum indicates label scrambling is not occurring.

4.3.4 Prevalence of Label Scrambling

To test the pervasiveness of label scrambling, we examined several other labeled peptides, including ones from a proteolytic digest of the protein β-2-microglobulin that had been labeled with DEPC for 1 minute. Each collection of peptides was separated by LC and then subjected to both CID and ETD. In all, 34 different labeled sites were probed (Table 4-1). The criteria for concluding that label scrambling had occurred were: (i) unlabeled and labeled versions of product ions were observed during CID and (ii) CID and ETD identified different labeled sites in a given peptide. Of these 34 labeled

peptides, nine showed evidence for scrambling. Interestingly, all nine had a histidine residue as either the originally labeled site or the site to which the labeled was transferred, indicating the lability of labeled histidine residues. Another interesting observation from this limited data set is that an acidic residue is present in all the peptides that undergo scrambling. Because of the effect of proton mobility on the prevalence of this scrambling reaction (see Figure 4-5), it is possible that intramolecular interactions between charged groups may be necessary to facilitate this reaction. Further work is clearly needed, though, to more fully understand how amino acid sequence affects this scrambling reaction.

Table 4-1 DEPC-labeled sites on peptides as identified by ETD and CID and whether scrambling has occurred.

Sequence	ETD of $(M+3H)^{3+}$	CID of $(M+3H)^{3+}$	Scrambling in $(M+3H)^{3+}$?	CID of $(M+2H)^{2+}$	Scrambling in $(M+2H)^{2+}$?
DRVYIHPFHL peak 1	Y4	Y4	No	Y4	No
DRVYIHPFHL peak 2	N-term	N-term	No	N-Term	No
DRVYIHPFHL peak 3	H6	H6	No	H6, H9	**Yes**
DRVYIHPFHL peak 4	H6	H6	No	H6, H9	**Yes**
GNHWAVGHLM-NH_2 peak 1	N-Term	N-Term	No	N-Term	No
GNHWAVGHLM-NH_2 peak 2	H8	H8	No	H8	No
GNHWAVGHLM-NH_2 peak 3	H3	H3	No	H3	No
GNHWAVGHLM-NH_2 peak 4	H8	H8	No	H8	No
GNHWAVGHLM-NH_2 peak 5	H3	H3	No	H3	No
DRVYIHPFHLVIH peak 1	Y4		No	Y4	No
DRVYIHPFHLVIH peak 2	N-term		No	N-term	No
DRVYIHPFHLVIH peak 3	H9		No	H6, H9, H13	**Yes**
DRVYIHPFHLVIH peak 4	H9		No	H6, H9	**Yes**
DRVYIHPFHLVIH peak 5	H13		No	H9, H13	**Yes**
DRVYIHPFHLVIH peak 6	H6		No	H6, H9	**Yes**
SYSMEHFRWGKPV peak 1	N-term	N-term	No	N-term, H6, K11	**Yes**
SYSMEHFRWGKPV peak 2	H6	H6	No	H6, K11	**Yes**
YEVHHQKLVFF peak 1	H4	H4, H5	**Yes**	H4, H5	**Yes**
YEVHHQKLVFF peak 2	N-term	N-Term	No	N-term	No
YEVHHQKLVFF peak 3	H4	H4	No	H4	No
YEVHHQKLVFF peak 4	H5	H5	No	H5	No
YEVHHQKLVFF peak 5	K7	K7	No	K7	No
SPWTKCSATCGGGHYMRTR peak 1	Y15		No	Y15/H14*	No
SPWTKCSATCGGGHYMRTR peak 2	H14		No	H14	No
SPWTKCSATCGGGHYMRTR peak 3	Y15/H14*		No	Y15/H14*	No
SPWTKCSATCGGGHYMRTR peak 4	N-term		No	N-Term	No
SPWTKCSATCGGGHYMRTR peak 5	K5		No	K5	No
SRHPAENGKSNF	S1/R2		No	N-term	No
SRHPAENGKSNF	H3		No	H3	No
IQRTPKIQVY	K6		No	Y3	No
IQRTPKIQVY	N-term		No	N-term	No
KNGERIEKVEHSDL	K1		No	K1	No
SQPKIVKW	K4		No	K4	No
Glp-HWSHGWYPG-NH_2 peak1	N-term	N-term	No	N-term	No
Glp-HWSHGWYPG-NH_2 peak2	His5	His5	No	His5	No
Glp-HWSHGWYPG-NH_2 peak3	His2	His2	No	His2	No
Glp-HWSHGWYPG-NH_2 peak4	His5	His5	No	His5	No
SRHPAENGKSNFLNCY	C15		No	C15	No

* Tandem mass spectrometry is unable to identify which of the listed residues is labeled due to the absence of the appropriate product ions.

4.4 Conclusions

Covalent labeling is increasingly used to study the surface structure of proteins and protein complexes, and DEPC is a labeling reagent that has shown great promise in these experiments because of its ability to label multiple residues simultaneously. Identifying the labeled sites is often done using CID after proteolytic digestion of the protein, but in this work, we show that scrambling of the DEPC label to another site on a peptide can occasionally occur during CID, resulting in ambiguous labeling site assignments. Like previous studies that reported the scrambling of functional groups (e.g. phosphate groups) on modified peptides, this scrambling occurs most readily under low proton mobility conditions. Also, scrambling does not appear to occur when ETD is used to dissociate the labeled peptides; this is consistent with the known mechanistic differences between CID and ETD. Based on the analysis of over 30 labeled peptides, we find evidence for scrambling about 25% of the time. In all the cases where scrambling is observed, histidine residues are involved. The increased tendency of histidine residues to undergo scrambling has important consequences as this residue is the most reactive amino acid with DEPC. A solution to this problem may be to produce larger peptides upon proteolysis so that the labeled peptides have higher charge states and are thus more amenable to ETD. Alternatively, supercharging agents could be used to produce higher peptide charge states to avoid the low proton mobility conditions that foster label scrambling during CID.

CHAPTER 5
SUMMARY AND FUTURE DIRECTIONS

5.1 Summary

This dissertation has investigated the application of MS-based techniques as a means to study protein structure. These techniques involve the use of chemical modifications to label solvent exposed sites. These modified sites are then measured via mass spectrometry. The location and frequency of these modifications are indicative of solvent accessibility or, in some cases, the dynamics. These values and, more importantly, changes in these values give insight into the tertiary structure and changes in tertiary structure near each label. These MS-based techniques have proven to be a powerful tool.

These methods were first applied to the amyloid forming protein β2m. This work illustrates that the binding of Cu(II) alone is capable of inducing the transformation of β2m into an amyloid competent state. Thus, explaining its unique ability to induce amyloid competent aggregation. In order to perform this, NMR and crystallographic structures of amyloidogenic conformers induced by other conditions were investigated in order to find common structural attributes. Two probable structural changes were found: (1) the burial of Trp60 and (2) the fraying of the B-, D-, and E- strands. All three complexes were probed looking for these changes. As expected, Cu(II) binding induces both of these changes. Ni(II) and Zn(II) however, did not. Suggesting that indeed, Cu(II) is

the only metal that induces the conformational change into the amyloidogenic conformer.

This research also potentially explained the amorphous aggregation of β2m upon the binding of Zn(II). This binding was shown to increase the dynamics of β2m in multiple strands of the protein. This increase in dynamics may be indicative of structural instability, which could then lead to aggregation. This is supported by the fact that Zn(II) binding has been shown to destabilize the structure of β2m by as much as 14 kJ/mol.[44] Two other conclusions that can be drawn from the work is that metal binding proximal to Pro32 and His31 alone is not sufficient to induce the conversion into the amyloidogenic conformer and that the destabilization of the D-strand alone is not enough to induce aggregation.

We then investigated the ability of a DEPC-based covalent labeling technique as a means to monitor the structure of therapeutic proteins. This was accomplished through the study of three proteins, β2m, IgG, and HGH. This work demonstrated that DEPC was able to identify structural perturbations after exposure to degradation conditions. These amino acid specific techniques require the simplest analysis and in the case of DEPC can probe up to 30% of the residues in a given protein, thus, providing a high degree of structural resolution when compared to other residue specific covalent labeling reagents. DEPC labeling is particularly valuable for identifying the interfaces of protein aggregates. In the case of IgG for example, the interface of heat induced aggregates was shown to be on the variable domains of the light and heavy chain. DEPC

labeling is also able to distinguish relatively minor perturbations in protein structure as illustrated by the experiments with HGH.

In order to identify the location of these covalent labels in the experiments above, the protein is subjected to proteolysis by trypsin. The resulting proteolytic fragments are then subjected to collision induced dissociation and the resulting tandem mass spectra can be used to identify the label. In this work however, we show that these DEPC labels can scramble to another site on a peptide during CID, resulting in ambiguous labeling site assignments. Similarly to the scrambling of other functional groups (e.g. phosphate groups) on modified peptides, this scrambling occurs most readily under low proton mobility conditions. This scrambling does not appear to occur when ETD is used to dissociate the labeled peptides; this is consistent with the known mechanistic differences between CID and ETD. Based on the analysis of over 30 labeled peptides, we find evidence for scrambling about 25% of the time. In all the cases where scrambling is observed, histidine residues are involved. The increased tendency of histidine residues to undergo scrambling has important consequences as this residue is the most reactive amino acid with DEPC. A solution to this problem may be to produce larger peptides upon proteolysis so that the labeled peptides have higher charge states and are thus more amenable to ETD. Alternatively, supercharging agents could be used to produce higher peptide charge states to avoid the low proton mobility conditions that foster label scrambling during CID.

5.2 Future directions

The following sections will discuss the potential applications, characterizations, and improvements for MS-spectrometry based structural techniques. This will also further discuss the mechanisms of DEPC scrambling and β2m aggregation.

5.2.1 The applications of Trp60 labeling

The ability to easily monitor the conversion of β2m into the amyloidogenic conformer is a powerful tool. It could be utilized to probe all known conditions shown to induce β2m oligomerization. Cataloging which conditions promote a mechanism similar to that of Cu(II) binding could shed light on the amyloidosis of proteins as a whole.

There are several small molecules (i.e. Epigallocatechin-3-gallate and suramin) that have been shown to inhibit or modulate the formation of fibrils. The mechanisms by which these drugs perform this modulation are unknown. This labeling technique could be used to identify if these molecules act on the amyloidogenic conformer or prevent its formation.

Previous work has shown that ΔN6 is capable of transforming natively folded β2m into amyloids. HNSB labeling could potentially be a way to monitor this transformation and potentially measuring its kinetics. As of now, to our knowledge, no other method capable of probing the kinetics of this reaction has been developed.

5.2.2 Investigating the sensitivity of DEPC labeling

Protein therapeutics are rapidly transforming the pharmaceutical industry. Current technologies are challenged to provide the rapid, high resolution analyses of protein higher order structures needed to ensure drug efficacy and safety. In order for DEPC labeling to become widely applicable in this field it must be sufficiently characterized against currently used methodology. Its ability to measure subtle structural changes upon mishandling is of particular interest.

5.2.3 Mechanism of Scrambling

Covalent labeling along with mass spectrometry is finding more use as a means of studying the higher order structure of proteins and protein complexes. Diethylpyrocarbonate (DEPC) is an increasingly used reagent for these labeling experiments because it is capable of modifying multiple residues at the same time. As this technique grows in use it is important to understand the mechanism behind this scrambling. Thus, investigating the effect of distance and chemistry on this scrambling is vital. Additionally, investigating the influence of acids or lack thereof on this scrambling could also be enlightening.

BIBLIOGRAPHY

1. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181,** 223–230 (1973).

2. Tanford, C. How protein chemists learned about the hydrophobic factor. *Protein Sci.* **6,** 1358–1366 (1997).

3. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6233–7 (1986).

4. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci* **101,** 9528–9533 (2004).

5. Standing, K. Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol.* **13,** 595–601 (2003).

6. Steen, H. & Mann, M. The ABC’s (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5,** 699–711 (2004).

7. Kelly, S. & Price, N. The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Curr. Protein Pept. Sci.* **1,** 349–384 (2000).

8. Surewicz, W. K., Mantsch, H. H. & Chapman, D. Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment. *Biochemistry* **32,** 389–394 (1993).

9. Wagner, G., Hyberts, S. G. & Havel, T. F. {NMR} structure determination in solution: a critique and comparison with x-ray crystallography. *Ann. Rev. Biophys. Biomol. Struct.* **21,** 167–198 (1992).

10. Stryer, L. Implications of X-ray crystallographic studies of protein structure. *Annu Rev Biochem* **37,** 25–50 (1968).

11. Chen, B., Retzlaff, M., Roos, T. & Frydman, J. Cellular strategies of protein quality control. *Cold Spring Harb. Perspect. Biol.* **3,** 1–14 (2011).

12. Huang, L., Liu, X., Cheng, B. & Huang, K. How our bodies fight amyloidosis: Effects of physiological factors on pathogenic aggregation of amyloidogenic proteins. *Arch. Biochem. Biophys.* **568,** 46–55 (2015).

13. Zhuravlev, P. I., Reddy, G., Straub, J. E. & D Thirumalai. Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *J. Mol. Biol.* (2014). doi:10.1016/j.jmb.2014.05.007

14. Ono, K. *et al.* In Vitro Amyloidogenic Peptides of Galectin-7: POSSIBLE MECHANISM OF AMYLOIDOGENESIS OF PRIMARY LOCALIZED CUTANEOUS AMYLOIDOSIS. *J. Biol. Chem.* **289,** 29195–207 (2014).

15. Tipping, K. W. *et al.* pH-induced molecular shedding drives the formation of amyloid fibril-derived oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 5691–5696 (2015).

16. Myers, S. L. *et al.* A systematic study of the effect of physiological factors on beta2-microglobulin amyloid formation at neutral pH. *Biochemistry* **45,** 2311–21 (2006).

17. Chiti, F. *et al.* A partially structured species of beta 2-microglobulin is significantly populated under physiological conditions and involved in fibrillogenesis. *J. Biol. Chem.* **276,** 46714–21 (2001).

18. Bush, A. I. & Tanzi, R. E. The galvanization of β-amyloid in Alzheimer's disease. *Proc Natl Acad Sci* **99,** 7317–7319 (2002).

19. Jobling, M. F. *et al.* Copper and zinc binding modulates the aggregation and neurotoxic properties of the prion peptide PrP106-126. *Biochemistry* **40,** 8073–84 (2001).

20. Uversky, V. N., Li, J. & Fink, a L. Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson's disease and heavy metal exposure. *J. Biol. Chem.* **276,** 44284–96 (2001).

21. Moriarty, G. M., Minetti, C. a S. a, Remeta, D. P. & Baum, J. A Revised Picture of the Cu(II)-α-Synuclein Complex: The Role of N-Terminal Acetylation. *Biochemistry* **53,** 2815–7 (2014).

22. Dong, J. *et al.* Unique effect of Cu(II) in the metal-induced amyloid formation of β-2-microglobulin. *Biochemistry* **53,** 1263–74 (2014).

23. Zhang, H., Rochet, J.-C. & Stanciu, L. A. Cu(II) promotes amyloid pore formation. *Biochem. Biophys. Res. Commun.* **464,** 342–7 (2015).

24. Dong, J. *et al.* Engineering metal ion coordination to regulate amyloid fibril assembly and toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 13313–8 (2007).

25. Tõugu, V. *et al.* Zn(II)- and Cu(II)-induced non-fibrillar aggregates of amyloid-beta (1-42) peptide are transformed to amyloid fibrils, both spontaneously and under the influence of metal chelators. *J. Neurochem.* **110,** 1784–95 (2009).

26. Miura, T., Suzuki, K., Kohata, N. & Takeuchi, H. Metal Binding Modes of Alzheimer's Amyloid β-Peptide in Insoluble Aggregates and Soluble Complexes †. *Biochemistry* **39,** 7024–7031 (2000).

27. Rowinska-Zyrek, M., Salerno, M. & Kozlowski, H. Neurodegenerative diseases - understanding their molecular bases and progress in the development of potential treatments. *Coord. Chem. Rev.* (2014). doi:10.1016/j.ccr.2014.03.026

28. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **70,** 1035–48 (1992).

29. Trinh, C. H., Smith, D. P., Kalverda, A. P., Phillips, S. E. V & Radford, S. E. Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9771–6 (2002).

30. Connors, L. H., Shirahama, T., Skinner, M., Fenves, A. & Cohen, A. S. Formation of Amyloid Fibrils From Intact [beta]2-Microglobulin. *Biochem. Biophys. Res. Commun.* **131,** 1063–1068 (1985).

31. Gejyo, F. *et al.* Beta 2-microglobulin: a new form of amyloid protein associated

with chronic hemodialysis. *Kidney Int.* **30,** 385–90 (1986).

32. Ayers, D. C., Athanasou, N. a, Woods, C. G. & Duthie, R. B. Dialysis arthropathy of the hip. *Clin. Orthop. Relat. Res.* 216–24 (1993). at <http://www.ncbi.nlm.nih.gov/pubmed/8472452>

33. Menaa, C., Esser, E. & Sprague, S. M. β2-Microglobulin stimulates osteoclast formation. *Kidney Int.* **73,** 1275–1281 (2008).

34. Floege, J. & Ketteler, M. beta2-Microglobulin-derived amyloidosis: An update. *Kidney Int.* **59,** 164–171 (2001).

35. Keating, M. J. Chronic lymphocytic leukemia. *Semin. Oncol.* **26,** 107–14 (1999).

36. Malaguarnera, M. *et al.* No Title. *Dig. Dis. Sci.* **42,** 762–766 (1997).

37. McParland, V. J. *et al.* Partially Unfolded States of β 2 -Microglobulin and Amyloid Formation in Vitro †. *Biochemistry* **39,** 8735–8746 (2000).

38. Naiki, H. *et al.* Establishment of a kinetic model of dialysis-related amyloid fibril extension in vitro. *Amyloid* **4,** 223–232 (1997).

39. Eichner, T. & Radford, S. E. A Generic Mechanism of β2-Microglobulin Amyloid Assembly at Neutral pH Involving a Specific Proline Switch. *J. Mol. Biol.* **386,** 1312–1326 (2009).

40. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. & Radford, S. E. A Systematic Investigation into the Effect of Protein Destabilisation on Beta 2-Microglobulin Amyloid Formation. *J. Mol. Biol.* **330,** 943–954 (2003).

41. Esposito, G. *et al.* Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci.* **9,** 831–45 (2000).

42. Heegaard, N. H. H., Roepstorff, P., Melberg, S. G. & Nissen, M. H. Cleaved beta 2-microglobulin partially attains a conformation that has amyloidogenic features. *J. Biol. Chem.* **277,** 11184–9 (2002).

43. Morgan, C. J., Gelfand, M., Atreya, C. & Miranker, a D. Kidney dialysis-associated amyloidosis: a molecular role for copper in fiber formation. *J. Mol. Biol.* **309,** 339–45 (2001).

44. Eakin, C. M., Knight, J. D., Morgan, C. J., Gelfand, M. A. & Miranker, A. D. Formation of a copper specific binding site in non-native states of beta-2-microglobulin. *Biochemistry* **41,** 10646–56 (2002).

45. Kumar, S., Sharma, P., Arora, K., Raje, M. & Guptasarma, P. Calcium binding to Beta-2-microglobulin at physiological ph drives the occurrence of conformational changes which cause the protein to precipitate into amorphous forms that subsequently transform into amyloid aggregates. *PLoS One* **9,** e95725 (2014).

46. Blaho, D. V & Miranker, A. D. Delineating the conformational elements responsible for Cu(2+)-induced oligomerization of beta-2 microglobulin. *Biochemistry* **48,** 6610–7 (2009).

47. Calabrese, M. F. & Miranker, A. D. Metal binding sheds light on mechanisms of amyloid assembly. *Prion* **3,** 1–4 (2009).

48. Yamamoto, S. Glycosaminoglycans Enhance the Trifluoroethanol-Induced Extension of 2-Microglobulin-Related Amyloid Fibrils at a Neutral pH. *J. Am. Soc. Nephrol.* **15,** 126–133 (2004).

49. Relini, A. *et al.* Collagen plays an active role in the aggregation of beta2-microglobulin under physiopathological conditions of dialysis-related amyloidosis. *J. Biol. Chem.* **281,** 16521–9 (2006).

50. McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. Structural properties of an amyloid precursor of beta(2)-microglobulin. *Nat. Struct. Biol.* **9,** 326–31 (2002).

51. Miura, Y. *et al.* Radiolucent bone cysts and the type of dialysis membrane used in patients undergoing long-term hemodialysis. *Nephron* **60,** 268–73 (1992).

52. van Ypersele de Strihou, C., Jadoul, M., Malghem, J., Maldague, B. & Jamart, J. Effect of dialysis membrane and patient's age on signs of dialysis-related amyloidosis. The Working Party on Dialysis Amyloidosis. *Kidney Int.* **39,** 1012–9 (1991).

53. Eakin, C. M., Attenello, F. J., Morgan, C. J. & Miranker, A. D. Oligomeric assembly of native-like precursors precedes amyloid formation by beta-2 microglobulin. *Biochemistry* **43,** 7808–15 (2004).

54. Mendoza, V. L., Antwi, K., Barón-Rodríguez, M. a, Blanco, C. & Vachet, R. W. Structure of the preamyloid dimer of beta-2-microglobulin from covalent labeling and mass spectrometry. *Biochemistry* **49,** 1522–32 (2010).

55. Srikanth, R., Mendoza, V. L., Bridgewater, J. D., Zhang, G. & Vachet, R. W. Copper binding to beta-2-microglobulin and its pre-amyloid oligomers. *Biochemistry* **48,** 9871–81 (2009).

56. Mendoza, V. L., Barón-Rodríguez, M. A., Blanco, C. & Vachet, R. W. Structural insights into the pre-amyloid tetramer of β-2-microglobulin from covalent labeling and mass spectrometry. *Biochemistry* **50,** 6711–22 (2011).

57. Antwi, K. *et al.* Cu(II) organizes β-2-microglobulin oligomers but is released upon amyloid formation. *Protein Sci.* **17,** 748–759 (2008).

58. Calabrese, M. F. & Miranker, A. D. Formation of a stable oligomer of beta-2 microglobulin requires only transient encounter with Cu(II). *J. Mol. Biol.* **367,** 1–7 (2007).

59. Villanueva, J. *et al.* Increase in the conformational flexibility of beta 2-microglobulin upon copper binding: a possible role for copper in dialysis-related amyloidosis. *Protein Sci.* **13,** 797–809 (2004).

60. De Lorenzi, E. *et al.* Capillary electrophoresis investigation of a partially unfolded conformation of beta(2)-microglobulin. *Electrophoresis* **23,** 918–25 (2002).

61. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41,** 161–72 (2011).

62. Jahn, T. R., Parker, M. J., Homans, S. W. & Radford, S. E. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* **13,** 195–201 (2006).

63. Eakin, C. M., Berman, A. J. & Miranker, A. D. A native to amyloidogenic transition regulated by a backbone trigger. *Nat. Struct. Mol. Biol.* **13,** 202–8 (2006).

64. Calabrese, M. F., Eakin, C. M., Wang, J. M. & Miranker, A. D. A regulatable switch mediates self-association in an immunoglobulin fold. *Nat. Struct. Mol. Biol.* **15,** 965–971 (2008).

65. Mendoza, V. L. & Vachet, R. W. Protein surface mapping using diethylpyrocarbonate with mass spectrometric detection. *Anal. Chem.* **80,** 2895–904 (2008).

66. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7,** 21–39 (2008).

67. Clark, A. . *et al.* BIOSYNTHETIC HUMAN INSULIN IN THE TREATMENT OF DIABETES. *Lancet* **320,** 354–357 (1982).

68. Keen, H. *et al.* HUMAN INSULIN PRODUCED BY RECOMBINANT DNA TECHNOLOGY: SAFETY AND HYPOGLYCÆMIC POTENCY IN HEALTHY MEN. *Lancet* **316,** 398–401 (1980).

69. Mason, A. S. Therapeutic use of human growth hormone. *J. Clin. Pathol. Suppl. (Assoc. Clin. Pathol).* **7,** 58–61 (1976).

70. Ayyar, V. History of growth hormone therapy. *Indian J. Endocrinol. Metab.* **15,** 162 (2011).

71. Dimitrov, D. S. Therapeutic Proteins. *Methods Mol. Biol.* **899,** 1–26 (2012).

72. Leavy, O. Therapeutic antibodies: past, present and future. *Nat. Rev. Immunol.* **10,** 297 (2010).

73. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: Successes, limitations and hopes for the future. *Br. J. Pharmacol.* **157,** 220–233 (2009).

74. Kaur, P. *et al.* Characterizing monoclonal antibody structure by carboxyl group

footprinting. *MAbs* **7,** 540–552 (2015).

75. Chirino, A. J., Ary, M. L. & Marshall, S. a. Minimizing the immunogenicity of protein therapeutics. *Drug Discov. Today* **9,** 82–90 (2004).

76. Veurink, M., Westermaier, Y., Gurny, R. & Scapozza, L. Breaking the aggregation of the monoclonal antibody bevacizumab (Avastin??) by dexamethasone phosphate: Insights from molecular modelling and asymmetrical flow field-flow fractionation. *Pharm. Res.* **30,** 1176–1187 (2013).

77. Van Buren, N., Rehder, D., Gadgil, H., Matsumura, M. & Jacob, J. Elucidation of two major aggregation pathways in an IgG2 antibody. *J. Pharm. Sci.* **98,** 3013–3030 (2009).

78. Chennamsetty, N., Voynov, V., Kayser, V., Helk, B. & Trout, B. L. Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11937–42 (2009).

79. Hawe, A., Friess, W., Sutter, M. & Jiskoot, W. Online fluorescent dye detection method for the characterization of immunoglobulin G aggregation by size exclusion chromatography and asymmetrical flow field flow fractionation. *Anal. Biochem.* **378,** 115–122 (2008).

80. Hawe, A., Kasper, J. C., Friess, W. & Jiskoot, W. Structural properties of monoclonal antibody aggregates induced by freeze-thawing and thermal stress. *Eur. J. Pharm. Sci.* **38,** 79–87 (2009).

81. Zandomeneghi, G., Krebs, M. R. H., McCammon, M. G. & Fändrich, M. FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Sci.* **13,** 3314–21 (2004).

82. Vermeer, a W. & Norde, W. The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. *Biophys. J.* **78,** 394–404 (2000).

83. Watson, C. & Sharp, J. S. Conformational analysis of therapeutic proteins by hydroxyl radical protein footprinting. *AAPS J.* **14,** 206–17 (2012).

84. Arakawa, T., Ejima, D., Li, T. & Philo, J. S. The critical role of mobile phase

composition in size exclusion chromatography of protein pharmaceuticals. *J. Pharm. Sci.* **99,** 1674–92 (2010).

85. Fincke, A., Winter, J., Bunte, T. & Olbrich, C. Thermally induced degradation pathways of three different antibody-based drug development candidates. *Eur. J. Pharm. Sci.* **62,** 148–160 (2014).

86. Hall, D. & Huang, L. On the use of size exclusion chromatography for the resolution of mixed amyloid aggregate distributions: I. Equilibrium partition models. *Anal. Biochem.* **426,** 69–85 (2012).

87. Pelaez-Aguilar, A. *et al.* Inhibition of light chain 6aJL2-R24G amyloid fiber formation associated with AL amyloidosis. *Biochemistry* 150727170821008 (2015). doi:10.1021/acs.biochem.5b00288

88. Schlenzig, D. *et al.* Pyroglutamate formation influences solubility and amyloidogenicity of amyloid peptides. *Biochemistry* **48,** 7072–7078 (2009).

89. Brummitt, R. K. *et al.* Nonnative aggregation of an IgG1 antibody in acidic conditions: part 1. Unfolding, colloidal interactions, and formation of high-molecular-weight aggregates. *J. Pharm. Sci.* **100,** 2087–103 (2011).

90. Joubert, M. K., Luo, Q., Nashed-Samuel, Y., Wypych, J. & Narhi, L. O. Classification and characterization of therapeutic antibody aggregates. *J. Biol. Chem.* **286,** 25118–25133 (2011).

91. Iwata, K. *et al.* 3D structure of amyloid protofilaments of beta2-microglobulin fragment probed by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 18119–24 (2006).

92. Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. Dynamics in the unfolded state of beta2-microglobulin studied by NMR. *J. Mol. Biol.* **346,** 279–94 (2005).

93. Rosenman, D. J., Connors, C. R., Chen, W., Wang, C. & García, A. E. Aβ monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425,** 3338–59 (2013).

94. Katou, H. *et al.* The role of disulfide bond in the amyloidogenic state of beta(2)-microglobulin studied by heteronuclear NMR. *Protein Sci.* **11,** 2218–29 (2002).

95. Clore, G. M., Martin, S. R. & Gronenborn, a M. Solution structure of human growth hormone releasing factor. Combined use of circular dichroism and nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **191,** 553–561 (1986).

96. Kasimova, M. R. *et al.* NMR studies of the backbone flexibility and structure of human growth hormone: a comparison of high and low pH conformations. *J. Mol. Biol.* **318,** 679–95 (2002).

97. Rosano, C., Zuccotti, S. & Bolognesi, M. The three-dimensional structure of beta2 microglobulin: results from X-ray crystallography. *Biochim. Biophys. Acta* **1753,** 85–91 (2005).

98. Iwata, K., Matsuura, T., Sakurai, K., Nakagawa, A. & Goto, Y. High-resolution crystal structure of beta2-microglobulin formed at pH 7.0. *J. Biochem.* **142,** 413–9 (2007).

99. Zuccotti, S., Rosano, C., Mangione, P., Bellotti, V. & Bolognesi, M. Preliminary crystallographic characterization of the human β2 microglobulin His31Tyr mutant in a tetrameric assembly. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 1270–1272 (2003).

100. Jion, A. I., Goh, L.-T. & Oh, S. K. W. Crystallization of IgG1 by mapping its liquid-liquid phase separation curves. *Biotechnol. Bioeng.* **95,** 911–8 (2006).

101. Harris, L. J., Skaletsky, E. & McPherson, A. Crystallization of intact monoclonal antibodies. *Proteins Struct. Funct. Genet.* **23,** 285–289 (1995).

102. Mirsky, I. A., Jinks, R. & Perisutti, G. The Isolation and Crystallization of Human Insulin. *J. Clin. Invest.* **42,** 1869–72 (1963).

103. Johnson, W. C. Protein secondary structure and circular dichroism: a practical guide. *Proteins* **7,** 205–14 (1990).

104. Jackson, M. & Mantsch, H. H. The Use and Misuse of FTIR Spectroscopy in the Determination of Protein Structure. *Crit. Rev. Biochem. Mol. Biol.* (2008). at

<http://www.tandfonline.com/doi/abs/10.3109/10409239509085140>

105. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1,** 2876–90 (2006).

106. Vigano, C., Manciu, L., Buyse, F., Goormaghtigh, E. & Ruysschaert, J. M. Attenuated total reflection IR spectroscopy as a tool to investigate the structure, orientation and tertiary structure changes in peptides and membrane proteins. *Biopolymers* **55,** 373–380 (2000).

107. Manavalan, P. & Johnson, W. C. Sensitivity of circular dichroism to protein tertiary structure class. *Nature* **305,** 831–832 (1983).

108. Royer, C. a. Probing Protein Folding and Conformational Transitions with Fluorescence Probing Protein Folding and Conformational Transitions with Fluorescence. **106,** 1769–1784 (2006).

109. Vallée-Bélisle, A. & Michnick, S. W. Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nat. Struct. Mol. Biol.* **19,** 731–6 (2012).

110. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm. Res.* **25,** 1487–99 (2008).

111. Kihara, M. *et al.* Conformation of amyloid fibrils of beta2-microglobulin probed by tryptophan mutagenesis. *J. Biol. Chem.* **281,** 31061–9 (2006).

112. Kameda, A. *et al.* Nuclear magnetic resonance characterization of the refolding intermediate of beta2-microglobulin trapped by non-native prolyl peptide bond. *J. Mol. Biol.* **348,** 383–97 (2005).

113. Sakata, M. *et al.* Kinetic coupling of folding and prolyl isomerization of beta2-microglobulin studied by mutational analysis. *J. Mol. Biol.* **382,** 1242–55 (2008).

114. Deperalta, G. *et al.* Structural analysis of a therapeutic monoclonal antibody dimer by hydroxyl radical footprinting. *MAbs* **5,** 86–101 (2012).

115. Sukumar, M., Storms, S. M. & De Felippis, M. R. Non-native intermediate

conformational states of human growth hormone in the presence of organic solvents. *Pharm. Res.* **22,** 789–796 (2005).

116. Khurana, R. *et al.* Mechanism of thioflavin T binding to amyloid fibrils. *J. Struct. Biol.* **151,** 229–238 (2005).

117. Wolfe, L. S. *et al.* Protein-induced photophysical changes to the amyloid indicator dye thioflavin T. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16863–8 (2010).

118. Yagi, H., Abe, Y., Takayanagi, N. & Goto, Y. Elongation of amyloid fibrils through lateral binding of monomers revealed by total internal reflection fluorescence microscopy. *Biochim. Biophys. Acta* **1844,** 1881–8 (2014).

119. Gohlke, J. R. the most popular fluorescence probes in use at this time . Weber. (1972).

120. Franey, H., Brych, S. R., Kolvenbach, C. G. & Rajan, R. S. Increased aggregation propensity of IgG2 subclass over IgG1: Role of conformational changes and covalent character in isolated aggregates. *Protein Sci.* **19,** 1601–1615 (2010).

121. Filipe, V. *et al.* Immunogenicity of different stressed IgG monoclonal antibody formulations in immune tolerant transgenic mice. *MAbs* **4,** 740–752 (2012).

122. Domanska, K. *et al.* Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic beta2-microglobulin variant. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 1314–9 (2011).

123. Ricagno, S., Raimondi, S., Giorgetti, S., Bellotti, V. & Bolognesi, M. Human beta-2 microglobulin W60V mutant structure: Implications for stability and amyloid aggregation. *Biochem. Biophys. Res. Commun.* **380,** 543–7 (2009).

124. Coales, S. J., Tuske, S. J., Tomasso, J. C. & Hamuro, Y. Epitope mapping by amide hydrogen/deuterium exchange coupled with immobilization of antibody, on-line proteolysis, liquid chromatography and mass spectrometry. *Rapid Commun. Mass Spectrom.* **23,** 639–47 (2009).

125. Latypov, R. F., Hogan, S., Lau, H., Gadgil, H. & Liu, D. Elucidation of acid-induced unfolding and aggregation of human immunoglobulin IgG1 and IgG2 Fc. *J. Biol.*

Chem. **287,** 1381–1396 (2012).

126. Clore, G. M. & Gronenborn, A. M. Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16,** 22–34 (1998).

127. Tzeng, S.-R., Pai, M.-T. & Kalodimos, C. G. NMR studies of large protein systems. *Methods Mol. Biol.* **831,** 133–40 (2012).

128. Sun, S. *et al.* Solid-state NMR spectroscopy of protein complexes. *Methods Mol. Biol.* **831,** 303–31 (2012).

129. Woodward, C., Simon, I. & Tuchsen, E. Hydrogen exchange and the dynamic structure of proteins. *Mol Cell Biochem* **48,** 135–160 (1982).

130. Kaltashov, I. A. & Eyles, S. J. Studies of biomolecular conformations and conformational dynamics by mass spectrometry. *Mass Spectrom. Rev.* **21,** 37–71 (2002).

131. Engen, J. R. & Smith, D. L. Peer Reviewed: Investigating Protein Structure and Dynamics by Hydrogen Exchange MS. *Anal. Chem.* **73,** 256 A–265 A (2001).

132. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M. & Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev. Proteomics* **8,** 43–59 (2011).

133. Hoofnagle, A. N., Resing, K. A. & Ahn, N. G. Protein analysis by hydrogen exchange mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* **32,** 1–25 (2003).

134. Percy, A. J., Rey, M., Burns, K. M. & Schriemer, D. C. Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. *Anal. Chim. Acta* **721,** 7–21 (2012).

135. Hodkinson, J. P., Jahn, T. R., Radford, S. E. & Ashcroft, A. E. HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein beta(2)-microglobulin upon release from the MHC-1. *J. Am. Soc. Mass Spectrom.* **20,** 278–86 (2009).

136. Hodkinson, J. P., Radford, S. E. & Ashcroft, A. E. The role of conformational

flexibility in β2-microglobulin amyloid fibril formation at neutral pH. *Rapid Commun. Mass Spectrom.* **26,** 1783–92 (2012).

137. Heegaard, N. H. H. *et al.* Unfolding, aggregation, and seeded amyloid formation of lysine-58-cleaved beta 2-microglobulin. *Biochemistry* **44,** 4397–407 (2005).

138. Beck, A. *et al.* Analytical characterization of biosimilar antibodies and Fc-fusion proteins. *TrAC Trends Anal. Chem.* **48,** 81–95 (2013).

139. Beck, A. *et al.* Cutting-edge mass spectrometry characterization of originator, biosimilar and biobetter antibodies. *J. Mass Spectrom.* **50,** 285–297 (2015).

140. Wei, H. *et al.* Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: Methodology and applications. *Drug Discov. Today* **19,** 95–102 (2014).

141. Zhang, A. *et al.* Understanding the conformational impact of chemical modifications on monoclonal antibodies with diverse sequence variation using hydrogen/deuterium exchange mass spectrometry and structural modeling. *Anal. Chem.* **86,** 3468–75 (2014).

142. Houde, D., Arndt, J., Domeier, W., Berkowitz, S. & Engen, J. R. Characterization of IgG1 conformation and conformational dynamics by hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **81,** 2644–51 (2009).

143. Zhang, Q. *et al.* Epitope mapping of a 95 kDa antigen in complex with antibody by solution-phase amide backbone hydrogen/deuterium exchange monitored by Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **83,** 7129–36 (2011).

144. Pandit, D. *et al.* Mapping of discontinuous conformational epitopes by amide hydrogen/deuterium exchange mass spectrometry and computational docking. *J. Mol. Recognit.* **25,** 114–24 (2012).

145. McCloskey, J. a. in *Methods Enzymol.* **193,** 329–338 (1990).

146. Hoerner, J. K., Xiao, H., Dobo, A. & Kaltashov, I. A. Is there hydrogen scrambling in the gas phase? Energetic and structural determinants of proton mobility within

protein ions. *J. Am. Chem. Soc.* **126,** 7709–17 (2004).

147. Demmers, J. A. A., Rijkers, D. T. S., Haverkamp, J., Killian, J. A. & Heck, A. J. R. Factors affecting gas-phase deuterium scrambling in peptide ions and their implications for protein structure determination. *J. Am. Chem. Soc.* **124,** 11191–8 (2002).

148. Hambly, D. M. & Gross, M. L. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J. Am. Soc. Mass Spectrom.* **16,** 2057–63 (2005).

149. Hambly, D. & Gross, M. Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin. *Int. J. Mass Spectrom.* **259,** 124–129 (2007).

150. Xu, G. & Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107,** 3514–43 (2007).

151. Konermann, L., Stocks, B. B., Pan, Y. & Tong, X. Mass spectrometry combined with oxidative labeling for exploring protein structure and folding. *Mass Spectrom. Rev.* **29,** 651–67 (2010).

152. Klinger, A. L. *et al.* A Synchrotron-Based Hydroxyl Radical Footprinting Analysis of Amyloid Fibrils and Prefibrillar Intermediates with Residue-Specific Resolution. *Biochemistry* (2014). doi:10.1021/bi5010409

153. Klinger, A. L., Kiselar, J., Chance, M. & Axelsen, P. H. Oxidative Footprinting of Fibrillar and Prefibrillar Oligomeric Forms of Amyloid Beta. *Biophys. J.* **104,** 399a (2013).

154. Klinger, A. L., Kiselar, J., Paravastu, A. & Rosenberry, T. Sequence Specific Radiolytic Footprinting Study of Monomer, Oligomeric and Fibrillar Amyloid Beta (1-42). *Biophys. J.* **108,** 495a (2015).

155. Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9,** 1634–49 (2010).

156. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional

protein structures and protein–protein interactions. *Mass Spectrom. Rev.* **25,** 663–682 (2006).

157. Sinz, A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **38,** 1225–1237 (2003).

158. Back, J. W., de Jong, L., Muijsers, A. O. & de Koster, C. G. Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **331,** 303–313 (2003).

159. Fancy, D. Elucidation of protein–protein interactions using chemical cross-linking or label transfer techniques. *Curr. Opin. Chem. Biol.* **4,** 28–33 (2000).

160. Kluger, R. & Alagic, A. Chemical cross-linking and protein–protein interactions—a review with illustrative protocols. *Bioorg. Chem.* **32,** 451–472 (2004).

161. Deroo, S. *et al.* Chemical Cross-Linking/Mass Spectrometry Maps the Amyloid β Peptide Binding Region on Both Apolipoprotein E Domains. *ACS Chem. Biol.* 150121144241005 (2015). doi:10.1021/cb500994j

162. Bitan, G. & Teplow, D. B. Rapid photochemical cross-linking--a new tool for studies of metastable, amyloidogenic protein assemblies. *Acc. Chem. Res.* **37,** 357–64 (2004).

163. Peter, J. F. & Tomer, K. B. A General Strategy for Epitope Mapping by Direct MALDI-TOF Mass Spectrometry Using Secondary Antibodies and Cross-Linking. *Anal. Chem.* **73,** 4012–4019 (2001).

164. Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **28,** 785–815 (2009).

165. Zhou, Y. & Vachet, R. W. Increased protein structural resolution from diethylpyrocarbonate-based covalent labeling and mass spectrometric detection. *J. Am. Soc. Mass Spectrom.* **23,** 708–17 (2012).

166. Kaur, P. *et al.* Characterizing monoclonal antibody structure by carbodiimide/GEE

footprinting. *MAbs* **6,** 1486–1499 (2014).

167. Zhou, Y. & Vachet, R. W. Covalent labeling with isotopically encoded reagents for faster structural analysis of proteins by mass spectrometry. *Anal. Chem.* **85,** 9664–70 (2013).

168. Fraczkiewicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319–333 (1998).

169. Xu, Y., Strickland, E. C. & Fitzgerald, M. C. Thermodynamic analysis of protein folding and stability using a tryptophan modification protocol. *Anal. Chem.* **86,** 7041–8 (2014).

170. Horton, H. R. & Koshland, D. E. A Highly Reactive Colored Reagent with Selectivity for the Tryptophan Residue in Proteins . **2,** (1965).

171. Aitken, M. *The Global Use of Medicines : Outlook Through 2015*. (2013). at <http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CFIQFjAA&url=http://www.imshealth.com/deployedfiles/ims/Global/Content/Insights/IMS Institute for Healthcare Informatics/Global_Use_of_Medicines_Report.pdf&ei=6czUT8nYEKuX0QXVxdisBA&usg>

172. Katta, V. & Chait, B. T. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun Mass Spectrom* **5,** 214–217 (1991).

173. Chowdhury, S. K., Katta, V. & Chait, B. T. Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **112,** 9012–9013 (1990).

174. Pirrone, G. F., Iacob, R. E. & Engen, J. R. Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Anal. Chem.* **87,** 99–118 (2015).

175. Pan, L. Y., Salas-Solano, O. & Valliere-Douglass, J. F. Antibody Structural Integrity of Site-Specific Antibody-Drug Conjugates Investigated by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **87,** 5669–5676 (2015).

176. Zhou, Y. & Vachet, R. W. Diethylpyrocarbonate labeling for the structural analysis of proteins: label scrambling in solution and how to avoid it. *J. Am. Soc. Mass Spectrom.* **23,** 899–907 (2012).

177. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30,** 918–20 (2012).

178. Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A. & Martens, L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* **11,** 996–9 (2011).

179. Craig, R. & Beavis, R. C. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun. Mass Spectrom.* **17,** 2310–6 (2003).

180. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–7 (2004).

181. Dorfer, V. *et al.* MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* **13,** 3679–84 (2014).

182. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5,** 5277–5286 (2014).

183. Geer, L. Y. *et al.* Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **3,** 958–964 (2004).

184. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13,** 22–4 (2013).

185. Vaudel, M. *et al.* PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* **33,** 1–46 (2015).

186. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11,** 395–406 (2010).

187. Wiedemann, C., Bellstedt, P. & Gorlach, M. CAPITO--a web server-based analysis

and plotting tool for circular dichroism data. *Bioinformatics* **29,** 1750–1757 (2013).

188. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4,** 1–13 (2009).

189. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

190. Pawar, A. P. *et al.* Prediction of 'aggregation-prone' and 'aggregation- susceptible' regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350,** 379–392 (2005).

191. DuBay, K. F. *et al.* Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* **341,** 1317–1326 (2004).

192. Tartaglia, G. G. *et al.* Prediction of Aggregation-Prone Regions in Structured Proteins. *J. Mol. Biol.* **380,** 425–436 (2008).

193. Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37,** 1395–1401 (2008).

194. Andersen, C. B., Manno, M., Rischel, C., Thórólfsson, M. & Martorana, V. Aggregation of a multidomain protein: A coagulation mechanism governs aggregation of a model IgG1 antibody under weak thermal stress. *Protein Sci.* **19,** 279–290 (2010).

195. Van Buren, N., Rehder, D., Gadgil, H., Matsumura, M. & Jacob, J. Elucidation of two major aggregation pathways in an IgG2 antibody. *J. Pharm. Sci.* **98,** 3013–3030 (2009).

196. Ito, T. & Tsumoto, K. Effects of subclass change on the structural stability of chimeric, humanized, and human antibodies under thermal stress. *Protein Sci.* **22,** 1542–1551 (2013).

197. Sahin, E., Grillo, A. O., Perkins, M. D. & Roberts, C. J. Comparative effects of pH and ionic strength on protein-protein interactions, unfolding, and aggregation for

IgG1 antibodies. *J. Pharm. Sci.* **99,** 4830–48 (2010).

198. Wu, H., Kroe-Barrett, R., Singh, S., Robinson, A. S. & Roberts, C. J. Competing aggregation pathways for monoclonal antibodies. *FEBS Lett.* **588,** 936–941 (2014).

199. Mulinacci, F., A.H. Capelle, M., Gurny, R., F. Drake, A. & Arvinte, T. Stability of human growth hormone: Influence of methionine oxidation on thermal folding. *J. Pharm. Sci.* **100,** 451–463 (2011).

200. Gau, B. C., Sharp, J. S., Rempel, D. L. & Gross, M. L. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Anal. Chem.* **81,** 6563–71 (2009).

201. Chen, J., Rempel, D. L. & Gross, M. L. Temperature jump and fast photochemical oxidation probe submillisecond protein folding. *J. Am. Chem. Soc.* **132,** 15502–4 (2010).

202. Konermann, L. & Pan, Y. Exploring membrane protein structural features by oxidative labeling and mass spectrometry. *Expert Rev. Proteomics* **9,** 497–504 (2012).

203. Chen, J., Rempel, D. L., Gau, B. C. & Gross, M. L. Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. *J. Am. Chem. Soc.* **134,** 18724–31 (2012).

204. Palumbo, A. M. & Reid, G. E. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal. Chem.* **80,** 9735–47 (2008).

205. Cui, L. & Reid, G. E. Examining factors that influence erroneous phosphorylation site localization via competing fragmentation and rearrangement reactions during ion trap CID-MS/MS and -MS(3.). *Proteomics* **13,** 964–73 (2013).

206. Gonzalez-Sanchez, M.-B., Lanucara, F., Hardman, G. E. & Eyers, C. E. Gas-phase intermolecular phosphate transfer within a phosphohistidine phosphopeptide dimer. *Int. J. Mass Spectrom.* 1–7 (2014). doi:10.1016/j.ijms.2014.04.015

207. Xiong, L., Ping, L., Yuan, B. & Wang, Y. Methyl group migration during the fragmentation of singly charged ions of trimethyllysine-containing peptides: precaution of using MS/MS of singly charged ions for interrogating peptide methylation. *J. Am. Soc. Mass Spectrom.* **20,** 1172–81 (2009).

208. Fuchs, R. & Budzikiewicz, H. Rearrangement reactions in the electrospray ionization mass spectra of pyoverdins. *Int. J. Mass Spectrom.* **210-211,** 603–612 (2001).